

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

**INTEGRACIÓN DE SISTEMAS DE
CERTIFICACIÓN MEDIANTE
ONTOLOGÍAS, CERTIFICATE
SUPPLEMENT Y TÉCNICAS DE
INFORMATION RETRIEVAL**

Cinthya Andrea Acosta Sepúlveda

INFORME FINAL DEL PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO CIVIL EN INFORMÁTICA

Julio del 2007

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

**INTEGRACIÓN DE SISTEMAS DE
CERTIFICACIÓN MEDIANTE
ONTOLOGÍAS, CERTIFICATE
SUPPLEMENT Y TÉCNICAS DE
INFORMATION RETRIEVAL**

Cinthyia Andrea Acosta Sepúlveda

Profesor Guía: Claudio Cubillos Figueroa

Profesor Co-referente: Cristian Alexandru Rusu

Carrera: Ingeniería Civil en Informática

Julio 2007

Dedicatoria

Gracias a mi familia y amigos por su apoyo incondicional durante todo este proceso, a Dios por darme fortaleza cuando el camino se hacía pesado, y a la universidad por la formación y los recuerdos que me deja. Esta etapa que ya terminó se la dedico a mi madre por su dedicación y esfuerzo, a mis hermanos por la ayuda y consejos que cada uno me dio. Sin ustedes no sería la persona que hoy soy. Gracias

Resumen

Dado el creciente interés de los estudiantes europeos de nivel técnico por ampliar sus estudios en países distintos al de origen, se hace necesario tener un proceso estandarizado en cuanto al reconocimiento y homologación de habilidades y competencias.

Es por ello que nace la iniciativa de integrar sistemas de certificación a nivel técnico, para facilitar así la movilidad de estudiantes y profesionales de este sector académico.

Para lograr este objetivo se utilizarán Ontologías para estandarizar la información perteneciente a los distintos programas académicos. El Certificate Supplement se usará como meta-modelo que permitirá la integración de certificados y haciendo uso de las técnicas de Information Retrieval se podrá obtener información para determinar si dos programas pueden ser considerados similares.

Palabras Claves: Ontología, Information Retrieval, Sistemas de Certificación, Certificate Supplement.

Abstract

Due to the growing interest from European students at technical level for studying abroad of their origin country, it has become necessary to have a standardized process of recognition and certification of their skills and knowledge acquired.

Therefore, was born the initiative to integrate the systems for certification at technical level, which facilitates the mobility of students and professional from the academic sector.

To achieve this goal it is necessary the use of Ontologies as a way to standardize the information belonging to different academic programs. The Certificate Supplement will be used as a meta-model that allows the certificates' integration and making use of Information Retrieval techniques to be able to obtain information to determine if two programs could be considerate similar.

TABLA DE CONTENIDOS

CAPÍTULO 1	1
INTRODUCCIÓN	1
1.1 DEFINICIÓN DE OBJETIVOS	5
1.2 METODOLOGÍA DE TRABAJO.....	5
1.3 ESTRUCTURACIÓN	6
CAPÍTULO 2	7
MECANISMOS DE INTEGRACIÓN	7
2.1 INTRODUCCIÓN.....	7
2.2 DEFINICIÓN DE ONTOLOGÍA.....	7
2.3 MAPPING DE ONTOLOGÍAS.....	8
2.4 KNOWLEDGE SHARING	16
2.4.1 <i>Servicios Compartidos Vía Traducción Punto a Punto</i>	17
2.4.2 <i>Formato de Intercambio Neutral</i>	18
2.4.3 <i>Autorización Neutral</i>	19
2.4.4 <i>Ontologías para Compartir Conocimiento</i>	21
2.4.5 <i>Mediadores</i>	28
2.5 INFORMATION RETRIEVAL.....	31
2.5.1 <i>Algoritmo Rocchio</i>	36
2.5.2 <i>Clasificador TFIDF</i>	38
2.5.3 <i>Clasificador Naive Bayes</i>	39
2.5.4 <i>Clasificador PrTFIDF</i>	41
2.5.5 <i>Clasificación Automática de Documentos</i>	42
CAPÍTULO 3	47
INICIATIVAS DE INTEGRACIÓN DE LOS SISTEMAS DE CERTIFICACIÓN A NIVEL EUROPEO 47	
3.1 INTRODUCCIÓN.....	47
3.2 EUROPEAN QUALIFICATION FRAMEWORK (EQF)	48
3.3 PROGRAMA SÓCRATES ERASMUS	51
3.4 EUROPEAN CREDIT TRANSFER AND ACCUMULATION SYSTEM (ECTS).....	53
3.5 EUROPASS.....	55
3.5.1 <i>Currículum Vitae</i>	55
3.5.2 <i>Pasaporte de Lenguas Europass</i>	56

3.5.3	<i>Documento de Movilidad Europass</i>	56
3.5.4	<i>Certificate Supplement Europass</i>	57
3.5.5	<i>Suplemento Europass al Título Superior</i>	58
3.6	PROGRAMA LEONARDO DA VINCI.....	58
3.6.1	<i>Proyecto B.E.A.T.R.I.C.</i>	59
3.7	EUROPEAN CREDITS TRANSFERT SYSTEM FOR VOCATIONAL EDUCATION AND TRAINING (ECVET) ..	60
3.7.1	<i>Principios de ECVET</i>	61
3.7.2	<i>Framework de Referencia para ECVET</i>	62
CAPÍTULO 4		64
MODELO CONCEPTUAL DE INTEGRACIÓN		64
4.1	INTRODUCCIÓN.....	64
4.2	LENGUAJE DE FORMALIZACIÓN	64
4.2.1	<i>Descripción de Redes Semánticas</i>	65
4.2.2	<i>Notación Formal Basada Sobre Redes Semánticas</i>	66
4.2.2.1	<i>Generalización</i>	68
4.2.2.2	<i>Asociación</i>	68
4.2.2.3	<i>Agregación</i>	69
4.3	CERTIFICATE SUPPLEMENT.....	70
4.3.1	<i>Descripción de la Red Semántica del Certificate Supplement</i>	73
4.4	ITALIA.....	78
4.4.1	<i>Introducción</i>	78
4.4.2	<i>Certificación en Italia</i>	78
4.4.3	<i>Descripción del Certificado IFTS</i>	78
4.4.4	<i>Descripción de la Red Semántica del Certificado IFTS</i>	80
4.5	IRLANDA.....	89
4.5.1	<i>Introducción</i>	89
4.5.2	<i>Certificación en Irlanda</i>	89
4.5.3	<i>Descripción de la Red Semántica del Certificado FETAC</i>	92
4.6	HOLANDA	96
4.6.1	<i>Introducción</i>	96
4.6.2	<i>Certificación en Holanda</i>	96
4.6.3	<i>Descripción del Certificado Holandés</i>	97
4.6.4	<i>Descripción de de la Red Semántica del Certificado de Cualificación</i>	99
4.7	REGLAS DE MAPPING.....	104
CAPÍTULO 5		106
DESARROLLO DE LA APLICACIÓN ILUSTRATIVA		106

5.1	INTRODUCCIÓN.....	106
5.2	REQUERIMIENTOS DE LA APLICACIÓN ILUSTRATIVA.....	107
5.3	DIAGRAMAS DE CASOS DE USO	108
5.3.1	<i>Diagrama de Caso de Uso: Generar C.S.....</i>	<i>108</i>
5.3.2	<i>Diagrama de Caso de Uso: Mostrar Ranking de Similitud.....</i>	<i>109</i>
5.4	DIAGRAMAS DE SECUENCIA.....	110
5.4.1	<i>Diagrama de Secuencia: Mostrar C.S.....</i>	<i>110</i>
5.4.2	<i>Diagrama de Secuencia: Mostrar Ranking de Similitud.....</i>	<i>110</i>
5.5	MODELOS ENTIDAD RELACIÓN.....	111
5.6	DESCRIPCIÓN CASO ESTUDIO HOLANDA	114
5.7	DESCRIPCIÓN CASO DE ESTUDIO ITALIA.....	114
5.8	DESCRIPCIÓN DE INFORMATION RETRIEVAL APLICADO A LOS CASOS DE ESTUDIO	115
5.9	NAVEGABILIDAD DE LA APLICACIÓN.....	119
5.10	RESULTADOS Y TRABAJOS FUTUROS	124
CAPÍTULO 6.....		129
CONCLUSIÓN.....		129

INDICE DE FIGURAS

FIGURA 2.1. INTEGRACIÓN UTILIZANDO UNA ÚNICA ONTOLOGÍA.....	10
FIGURA 2.2. INTEGRACIÓN UTILIZANDO MÚLTIPLES ONTOLOGÍAS.....	10
FIGURA 2.3. INTEGRACIÓN UTILIZANDO ONTOLOGÍAS HÍBRIDAS.....	12
FIGURA 2.4. REPRESENTACIÓN ONTOLOGÍA 1 USADA POR EL AGENTE 1.....	13
FIGURA 2.5. REPRESENTACIÓN ONTOLOGÍA 2 USADA POR EL AGENTE 2.....	13
FIGURA 2.6. REPRESENTACIÓN DE EXPRESIONES PARA LA ONTOLOGÍA 1.....	15
FIGURA 2.7. SERVICIO COMPARTIDO VÍA TRADUCCIÓN PUNTO A PUNTO MODELO SIMPLIFICADO.....	18
FIGURA 2.8. SERVICIO COMPARTIDO VÍA TRADUCCIÓN PUNTO A PUNTO PROCESO COMPLETO.....	18
FIGURA 2.9. FORMATO DE INTERCAMBIO NEUTRAL.....	19
FIGURA 2.10. MODELO DE AUTORIZACIÓN NEUTRAL.....	20
FIGURA 2.11. AUTORIZACIÓN NEUTRAL DE ONTOLOGÍAS.....	23
FIGURA 2.12. AUTORIZACIÓN DE DATOS OPERACIONALES.....	24
FIGURA 2.13. COMUNICACIÓN ENTRE HUMANOS.....	25
FIGURA 2.14. ACCESO DE DATOS VÍA ONTOLOGÍA COMPARTIDA.....	25
FIGURA 2.15. ACCESO DE DATOS VÍA MAPPING DE ONTOLOGÍA.....	26
FIGURA 2.16. SERVICIOS COMPARTIDOS.....	27
FIGURA 2.17. BÚSQUEDA BASADA EN CONCEPTOS.....	28
FIGURA 2.18. REPRESENTACIÓN DE UN VECTOR PARA UN DOCUMENTO.....	37
FIGURA 2.19. REPRESENTACIÓN DE LA FASE DE APRENDIZAJE.....	43
FIGURA 2.20. REPRESENTACIÓN DE LA FASE DE CLASIFICACIÓN.....	43
FIGURA 2.21. REPRESENTACIÓN DE LA CLASIFICACIÓN DE DOCUMENTOS.....	44
FIGURA 2.22. REPRESENTACIÓN DE LA RECUPERACIÓN DE DOCUMENTOS.....	45
FIGURA 3.1. PROPUESTA DEL EQF.....	50
FIGURA 4.1. REPRESENTACIÓN DE UN NODO COMPLEJO Y VARIOS NODOS BÁSICOS.....	67
FIGURA 4.2. REPRESENTACIÓN FORMAL DE LA GENERALIZACIÓN.....	68
FIGURA 4.3. REPRESENTACIÓN FORMAL DE LAS RESTRICCIONES DE CARDINALIDAD EN UNA RED SEMÁNTICA.....	69
FIGURA 4.4. REPRESENTACIÓN FORMAL DE LA AGREGACIÓN DENTRO DE UNA RED SEMÁNTICA.....	70
FIGURA 4.5. FORMATO DEL CERTIFICATE SUPPLEMENT (HOJA 1).....	71
FIGURA 4.6. FORMATO DEL CERTIFICATE SUPPLEMENT (HOJA 2).....	72
FIGURA 4.7. RED SEMÁNTICA DEL CERTIFICATE SUPPLEMENT.....	77
FIGURA 4.8. FORMATO DEL CERTIFICADO IFTS (PÁGINA 1).....	79
FIGURA 4.9. FORMATO DEL CERTIFICADO IFTS (PÁGINA 2).....	80
FIGURA 4.10. RED SEMÁNTICA DEL CERTIFICADO IFTS.....	88
FIGURA 4.11. FORMATO DEL CERTIFICADO FETAC.....	91
FIGURA 4.12. REPRESENTACIÓN FORMAL DE LA RED SEMÁNTICA DEL CERTIFICADO FETAC.....	95

FIGURA 4.13. FORMATO DEL CERTIFICATE OF QUALIFICATION.	98
FIGURA 4.14. REPRESENTACIÓN FORMAL DE LA RED SEMÁNTICA DEL CERTIFICATE OF QUALIFICATION.	103
FIGURA 5.1. DIAGRAMA CASO USO: GENERAR C.S.	109
FIGURA 5.2. DIAGRAMA CASO USO: MOSTRAR RANKING DE SIMILITUD.	109
FIGURA 5.3. DIAGRAMA DE SECUENCIA: MOSTRAR C.S.	110
FIGURA 5.4. DIAGRAMA DE SECUENCIA: MOSTRAR RANKING DE SIMILITUD.	111
FIGURA 5.5. MODELO ENTIDAD RELACIÓN ITALIA.	112
FIGURA 5.6. MODELO ENTIDAD RELACIÓN HOLANDA.	113
FIGURA 5.7. PANTALLA DE INICIO.	119
FIGURA 5.8. SELECCIÓN DE OPCIONES.	120
FIGURA 5.9. COMPARAR AWARDING BODY.	120
FIGURA 5.10. COMPARAR DURACIÓN DEL PROGRAMA.	120
FIGURA 5.11. SELECCIÓN DE COMPETENCIA.	121
FIGURA 5.12. MOSTRAR RANKING DE SIMILITUD.	121
FIGURA 5.13. MOSTRAR DESCRIPCIÓN DE COMPETENCIAS.	122
FIGURA 5.14. SELECCIÓN DE CERTIFICADOS.	122
FIGURA 5.15. CERTIFICATE SUPPLEMENT HOLANDA.	123
FIGURA 5.16. COMPETENCIAS CON MAYOR PORCENTAJE DE SIMILITUD.	124
FIGURA 5.17. COMPARACIÓN DE DESCRIPCIÓN DE COMPETENCIAS CON MAYOR PORCENTAJE DE SIMILITUD.	125
FIGURA 5.18. COMPARACIÓN DE COMPETENCIAS CON MENOR PORCENTAJE DE SIMILITUD.	126
FIGURA 5.19. DESCRIPCIÓN DE COMPETENCIAS CON MENOR PORCENTAJE DE SIMILITUD.	127

LISTA DE TABLAS

TABLA 4.1. MAPPING ENTRE EL CERTIFICATE SUPPLEMENT Y LOS PAÍSES ITALIA Y HOLANDA.	104
TABLA 5.1. PARTE DE ARCHIVO GENERADO PARA LLENAR LA TABLA “CA_PALABRAS”.	115
TABLA 5.2. LISTADO DE ALGUNOS STOP WORDS.	116
TABLA 5.3. EJEMPLO DEL RESULTADO OBTENIDO LUEGO DE APLICAR STEMMING.	117
TABLA 5.4. DESCRIPCIÓN Y COMPARACIÓN DE COMPETENCIAS CON MAYOR VALOR DE SIMILITUD.	125
TABLA 5.5. DESCRIPCIÓN Y COMPARACIÓN DE COMPETENCIAS CON MENOR VALOR DE SIMILITUD.	127

Capítulo 1

Introducción

Actualmente se observa un creciente interés en los estudiantes europeos con algún grado académico por ampliar sus estudios en otros países miembros de la comunidad. Sin embargo, este proceso no está estandarizado en cuanto al reconocimiento y homologación de habilidades y conocimientos adquiridos. Surge entonces la necesidad de establecer una certificación transparente, respecto de las cualificaciones y competencias de estudiantes y profesionales con pretensiones de aprendizaje permanente, para facilitar así su movilidad a través de Europa contribuyendo con el desarrollo de una educación y formación de calidad.

Dentro de este contexto aparece el proyecto “European Union. Leonardo da Vinci B.E.A.T.R.I.C [1]” que pretende transparentar los certificados nacionales emitidos por los diferentes países participantes del proyecto: Italia, Francia, Irlanda, Holanda, Alemania, Rumania, Grecia y Eslovenia. Orientándose a la educación post secundaria no universitaria VET (Vocational Education Training), es decir, carreras técnicas (Institutos Profesionales y Centros de Formación Técnica).

Cada país miembro de la comunidad europea tiene la posibilidad de otorgar de manera opcional el Certificate Supplement (C.S) a estudiantes y trabajadores para favorecer su movilidad dentro de la Unión Europea.

El C.S corresponde a un certificado de formación académica, perteneciente al pasaporte europeo Europass [2], que complementa el título profesional otorgado a un estudiante que culmina su formación en alguna institución universitaria. Fue diseñado para facilitar la comprensión del título profesional a personas de otros países en términos de competencias adquiridas por el individuo a lo largo de su trayectoria académica.

El documento describe las habilidades y competencias adquiridas por el profesional, la gama de trabajos a los que puede acceder, el nombre y estado legal de la comisión a cargo de la entrega y acreditación del documento, el nivel del certificado dentro del framework europeo, las diferentes maneras de adquirirlo y todos los requerimientos y oportunidades de acceder al siguiente nivel educacional.

Sin embargo el C.S, no apoya la movilidad de estudiantes con formación no universitaria debido a la dificultad de cuantificar las competencias a nivel técnico, es decir, poder medir con claridad el saber hacer.

La iniciativa de realizar este proyecto resulta de las necesidades expuestas en los párrafos anteriores, que es integrar sistemas de certificación a nivel de competencias para facilitar la movilidad de estudiantes y profesionales con conocimientos técnicos. Como requerimiento del Proyecto B.E.A.T.R.I.C se utilizará el “Certificate Supplement” como documento base ya que tiene una estructura estandarizada que facilita la comparación de certificados, es decir, es un meta-modelo que asiste la integración de certificados.

Del conjunto de países que participan en el proyecto “B.E.A.T.R.I.C” se escogió estudiar los procesos de certificación académica en Italia, Holanda e Irlanda, ya que se encuentra disponible información suficiente para poder obtener una visión global de estos procesos.

Tanto los certificados emitidos por los países mencionados y el Certificate Supplement se encuentran en un formato de texto, por lo tanto, un primer aporte de este proyecto fue la formalización de éstos.

Otro requerimiento del proyecto es el lenguaje de formalización de certificados que será Redes Semánticas, debido que las personas que analizarán estos modelos no pertenecen al área informática sino que son personas vinculadas al área educacional, por lo tanto, éstos son más fáciles de entender por personas que no tienen una formación en esta área.

Para poder formalizar los modelos de los países escogidos como casos de estudio fue necesario estudiar el mecanismo de certificación de cada país, con el fin de lograr una visión global del ámbito y complejidad del proceso de comparación e integración, que implica la construcción de un meta modelo a partir de los modelos existentes, conectando ideas, conceptos, relaciones, etc. Luego de esto, fue posible evaluar técnicas que permitieran integrar cada uno de estos modelos para establecer un documento estándar que permita comprender y comparar certificados de diferentes países y modelos. Es decir, establecer un mapping entre modelos que sirva de ayuda para estandarizar la descripción de la información, con el propósito de integrar modelos que se encuentran formalizados a través de modelos semánticamente distintos.

Dentro del ámbito de investigación cubierto, con respecto a lo existente en informática para integrar sistemas semánticamente distintos se abarcaron tres grandes áreas donde la primera de ellas fue el análisis de diversas técnicas de Mapping de Ontologías, para continuar con Knowledge Sharing y finalmente técnicas de Information Retrieval.

La investigación de cada una de estas áreas se realizó con el propósito de encontrar una metodología que sirva de ayuda para que un grupo de personas ubicadas en distintos países de Europa logren medir el grado de conocimientos adquiridos por estudiantes de un país que desee continuar sus estudios en otro, en otras palabras, ver qué competencias en términos de conocimientos y habilidades ha adquirido un alumno de nivel técnico en su país de origen y que pueden ser homologadas en el país destino, dentro del framework de cualificaciones que tiene éste para la misma carrera.

El mapping de ontologías es la tarea de relacionar vocabulario de dos ontologías que comparten el mismo dominio. Como describe Gruber [12] una ontología es una especificación explícita de una conceptualización. Una conceptualización es un modelo abstracto, una visión simplificada de algún fenómeno del mundo real que se desea representar para un cierto propósito, es decir, son objetos, conceptos, y otras entidades que se asumen que existen dentro de un área de interés y las relaciones que se manejan entre ellas. Dentro de esta área se estudió la posibilidad de hacer un mapping entre modelos de bases de datos (H. Wache [13]), donde se utiliza una ontología global que proporciona un vocabulario compartido para la especificación semántica. Luego se hace una copia de cada una de las estructuras de la base de datos, se codifica a un lenguaje que posibilite el razonamiento automático y finalmente se hace la integración sobre las copias.

Otra técnica estudiada dentro de la misma área fue el Mapping Inter-Ontologías que permite una gran flexibilidad, pero falla en asegurar la preservación semántica, ya que los usuarios pueden definir mapping arbitrarios.

El problema de la utilización de mapping como única técnica de integración para este proyecto, es que sólo da una integración a nivel de clases no a nivel de instancias, por lo tanto, no logran una granularidad fina al momento de integrar y comparar competencias que es lo necesario para el proyecto.

La segunda área de investigación fue Knowledge Sharing, donde se analizó la técnica de Servicios Compartidos Vía Traducción Punto a Punto (M. Uschold [20]), en la cual un sistema pone sus servicios a disposición de otro. Para ello se requiere el uso de un traductor, ya que los conceptos y términos utilizados en los sistemas que intentan compartir conocimiento no son iguales. Por lo tanto, se basa en la utilización de un sistema experto que responde a consultas hechas por otro sistema para un tema determinado.

Otra técnica estudiada dentro de esta área fue la utilización de un Formato de Intercambio Neutral, en ella se emplea un formato neutral para comunicar más de dos sistemas, ya que la expresividad del formato neutral cubre los lenguajes de todos los lenguajes objetivos individuales.

Las técnicas mencionadas son aptas para contestar solicitudes de información a través de sistemas expertos, pero no es útil para medir el grado de similitud entre modelos que es lo buscado por este proyecto.

La última área analizada fue Information Retrieval (I.R) (Lewis [18]), que es una técnica que guía al usuario a aquellos documentos que lo ayudan a satisfacer sus necesidades de información. Consta de 4 fases, Indexación o categorización de texto que convierte documentos originales a vectores, Formulación de consultas que consiste en generar consultas que interpreten los sistemas de software de I.R, Comparación donde se comparan las consultas con los documentos almacenados y Feedback que se produce entre el usuario y el sistema, a través del cual se hacen modificaciones a la consulta para alcanzar resultados aceptables.

La primera fase es la más importante (T. Joachims [16]), ya que la categorización de texto es la agrupación de documentos en diferentes categorías o clases, en la cual se realiza una exploración léxica y un análisis morfológico a la palabra, para luego generar los vectores de espacios que contendrán los pesos o valores que determinan la importancia de una palabra en un documento. Los pesos de una palabra se determinan en función a distribuciones estadísticas de las palabras que se encuentran en la base de datos y en el texto.

Dentro del área de Information Retrieval es donde se encontró una solución para poder comparar modelos semánticamente distintos a nivel de competencias, ya que los algoritmos que aquí se utilizan buscan la ocurrencia de una palabra en un documento, lo que es de gran utilidad porque determinar cuántas veces se encuentra una palabra dentro de una oración, puede servir para posteriormente comparar el grado de similitud entre las instancias de competencias en distintos modelos.

Para implementar el llenado del vector de pesos se utilizó el algoritmo TFIDF (Term Frequency / Inverse Document Frequency) utilizado en técnicas de Information Retrieval descrito en T. Joachims [17], el cual analiza la frecuencia de un término en un documento.

Para comprobar la funcionalidad del algoritmo se seleccionó como casos de estudio una pareja de carreras técnicas correspondientes a Italia y Holanda, ya que los certificados a nivel técnico de estos países ya fueron formalizados a través de redes semánticas en una primera iteración del proyecto. El modo de comprobación será a través de una aplicación ilustrativa que en primera instancia entregue el certificado de cada país en el formato del C.S, para posteriormente poder hacer la comparación de ambos certificados aplicando las técnicas de Information Retrieval. La aplicación será implementada a través de PHP y MySQL para el almacenamiento de la información en la base de datos.

1.1 Definición de Objetivos

La meta fundamental perseguida consiste en desarrollar un modelo conceptual de integración basado en el Certificate Supplement que permita comparar certificados de estudio para algunos países de la Unión Europea, semánticamente distintos.

Para ello es necesario cumplir los siguientes Objetivos Específicos:

- Hacer un modelo formal de los diferentes certificados de formación técnica utilizando como documento base el Certificate Supplement.
- Llegar a un mecanismo de integración utilizando el Certificate Supplement y técnicas de Information Retrieval que permita comprender y comparar certificados de diferentes países y distintos modelos.
- Validar el modelo conceptual de integración a través de una aplicación ilustrativa que asista en la comparación, utilizando el algoritmo TFIDF perteneciente a Information Retrieval aplicado a dos programas académicos pertenecientes a los países de Italia y Holanda.

1.2 Metodología de Trabajo

La metodología de trabajo para este proyecto cubrió los siguientes aspectos:

- Analizar cada uno de los sistemas de certificación de los países participantes del proyecto B.E.A.T.R.I.C. para escoger los más adecuados como caso de estudio.
- Elegir el mecanismo de formalización para los certificados de los países elegidos OWL, entidad relación, UML, redes semánticas, etc.
- Elegir la herramienta de modelado para el mecanismo de formalización elegido.
- Estado del arte de las diferentes metodologías de integración existentes.
- Refinar los requerimientos de la aplicación ilustrativa que se utilizará para validar el mecanismo de integración.
- Implementación del mecanismo de Integración a través de la tecnología escogida.
- Migrar los casos de estudio a la estructura de la aplicación ilustrativa.
- Diseñar la Base de Datos.
- Implementar la Base de Datos.
- Migrar la información de los casos de estudio a la Base de datos.
- Armar los escenarios de pruebas para los pares de países escogidos, Italia–Holanda.

- Ajustar los parámetros del algoritmo a los pares de países escogidos Italia–Holanda para las pruebas.
- Análisis de los resultados obtenidos en las pruebas.

1.3 Estructuración

En el capítulo dos se describirán las distintas áreas de integración de sistemas cubiertas que corresponden a Mapping de Ontologías, Knowledge Sharing e Information Retrieval, explicándose cada una de las técnicas pertenecientes a cada área de estudio.

El capítulo tres cubre las iniciativas de integración de los sistemas de certificación a nivel europeo describiéndose el European Qualification Framework (EQF), European Credit Transfer and Accumulation System (ECTS), qué es el Europass y cuál es el conjunto de documentos que entrega, dando un especial énfasis a la descripción del Certificate Supplement y los programas de movilidad estudiantil a nivel universitarios existentes.

El capítulo cuatro se refiere a los modelos conceptuales de integración, correspondientes a los certificados de formación técnica entregados por los países usados como casos de estudio Italia, Holanda, Irlanda, además del Certificate Supplement. También describe el lenguaje de formalización usado para establecer los modelos de los respectivos certificados. Otro punto dentro de este capítulo son las reglas de mapping establecidas para poder realizar la integración de certificados.

El capítulo cinco muestra el desarrollo de la aplicación ilustrativa, incluyendo los requerimientos tanto funcionales como no funcionales, los diagramas de casos de uso y diagramas de secuencias, hasta llegar a los modelos entidad relación usados para sustentar la base de datos de la aplicación. También se encuentra la descripción de los casos de estudios usados para probar la utilidad del algoritmo TFIDF perteneciente a Information Retrieval como una técnica que asiste a la comparación de certificados, además de mostrar a través de interfaces la navegabilidad de la aplicación y como último punto se muestran los resultados y trabajos futuros.

El capítulo seis está dedicado a la conclusión del trabajo realizado.

Capítulo 2

Mecanismos de Integración

2.1 Introducción

En este capítulo se dará una descripción del estado del arte de lo que existe en informática para tratar el problema de integración de modelos semánticamente distintos. Para ello se investigó sobre el Mapping de Ontologías, Knowledge Sharing e Information Retrieval.

La investigación de cada una de estas áreas se realizó con el propósito de encontrar una metodología que sirva de ayuda para que un grupo de personas ubicados en distintos países de Europa logre medir el grado de conocimientos adquiridos por estudiantes de un país que deseen continuar sus estudios en otro, es decir, ver qué competencias en términos de conocimientos y habilidades ha adquirido un alumno de nivel técnico en su país de origen y que pueden ser homologadas en el país destino, dentro del framework de cualificaciones que tiene éste para la misma carrera.

Antes de comenzar a detallar cada una de las áreas de estudio antes mencionadas junto con sus respectivas técnicas es necesario definir claramente un concepto al que se hará referencia en reiteradas ocasiones a lo largo de este capítulo y se trata del concepto de Ontología.

2.2 Definición de Ontología

El término Ontología tiene un origen filosófico, donde es definida como la rama de la metafísica que estudia la naturaleza de la existencia, también dentro de su origen epistemológico hace referencia al conocimiento y el saber.

Pero desde el punto de vista informático como describe Gruber [12] una ontología es una especificación explícita de una conceptualización. Una conceptualización es un modelo abstracto, una visión simplificada de algún fenómeno del mundo real que se desea representar para un cierto propósito, es decir, son objetos, conceptos, y otras entidades que se asumen que existen dentro de un área de interés y las relaciones que se manejan entre ellas.

Para los sistemas de inteligencia artificial lo que existe es lo que puede ser representado. Cuando hablamos de ontologías como "sistemas de representación de conocimiento" debemos especificar a qué tipo de sistemas nos referimos.

En realidad, las ontologías se están empleando en todo tipo de aplicaciones informáticas en las que sea necesario definir concretamente el conjunto de entidades relevantes en el campo de aplicación determinado, así como las interacciones entre las mismas. Algunas ontologías se crean con el mero objetivo de alcanzar una comprensión del "*Universo Declarativo*" pertinente, ya que su creación impone una especificación muy detallada. El "*Universo Declarativo*" de una ontología es el conjunto de objetos que están representados en ella y sobre los cuales se puede hablar y razonar.

2.3 Mapping de Ontologías

El hablar de mapping de ontologías se refiere a una manera de realizar integración de información desde fuentes de datos heterogéneas.

Dentro de la integración de información según H. Wache [13] las ontologías pueden ser utilizadas para describir la semántica de las fuentes de información y hacer explícito su contenido. Por lo tanto, son útiles para la integración de fuentes de datos, porque identifican y asocian conceptos semánticamente correspondientes.

Para poder utilizar una ontología es necesario solucionar el problema de la interoperabilidad, es decir, que la información que será compartida no sólo necesita proporcionar accesibilidad completa a los datos, sino que también requiere que los datos accesados puedan ser procesados e interpretados por sistemas remotos.

Los problemas que pueden surgir debido a la heterogeneidad de los datos ya son conocidos dentro de la comunidad de sistemas de bases de datos distribuidas y corresponden a la heterogeneidad semántica y la heterogeneidad estructural.

La heterogeneidad semántica se describe en H. Wache [13] y ocurre cuando diferentes sistemas de información almacenan sus datos en diferentes estructuras. Por lo tanto, se consideran los contenidos de un ítem de información y su significado. Para alcanzar la interoperabilidad semántica, el significado de la información que será intercambiada debe ser comprendido por todos los sistemas que interactuarán.

En F. Wiesman [14] se identifican tres conflictos que impiden lograr la heterogeneidad semántica:

- **Conflicto Estructural** que ocurre cuando los conceptos parecen tener el mismo significado, pero en realidad difieren, es decir, tienen diferencias en su estructura semántica.
- **Conflictos de datos** que ocurren cuando se utilizan diferentes representaciones del mismo dato.
- **Conflictos de Nombre** que ocurre cuando se utilizan diferentes nombres para el mismo tipo de información o los mismos nombres para diferentes tipos de información. Un fenómeno frecuente es la presencia de homónimos y de sinónimos.

El uso de ontologías para la aplicación de conocimiento implícito y oculto es una alternativa posible para superar el problema de la heterogeneidad semántica.

A continuación se discute una aplicación importante de mapping dentro de lo que es integración de información analizadas desde H.Wache [13] y se trata del mapping entre las ontologías y la información que ellas describen.

El mapping de ontologías se asocia con la tarea de relacionar el vocabulario de dos ontologías que comparten el mismo dominio, es decir, mapear conceptos encontrados en una ontología con los conceptos que se encuentren en la otra. El mapping no se puede percibir como un modelo autónomo del mundo, más bien debe parecer como un pegamento que junta información de distinta clase.

La primera y más obvia aplicación de mapping es relacionar la ontología al contenido actual del sistema de información. Una ontología se puede relacionar a modelos de base de dato así como a un único término dentro de la base de datos.

A continuación se describen tres formas diferentes de integrar información a través de ontologías.

1. Utilizando una Única Ontología

Este enfoque utiliza una única ontología global que proporciona un vocabulario común para la especificación semántica. Todas las fuentes de información se relacionan a esta ontología global como se muestra en la figura 2.1.

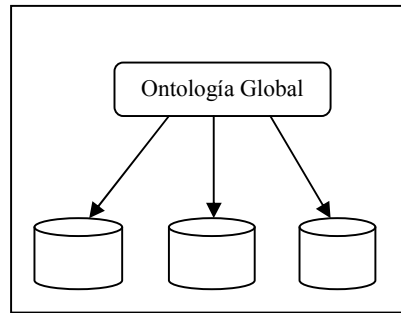


Figura 2.1. Integración utilizando una única ontología.

La ontología global también puede ser combinada con varias ontologías especializadas. Esto se puede realizar importando otros módulos de distintas ontologías.

Esta metodología de unificación se puede aplicar cuando todas las fuentes de información que serán integradas proporcionan casi la misma visión del dominio. Pero si una fuente de información tiene una visión distinta del dominio este tipo de mapping se vuelve dificultoso.

La aplicación de mapping a través de una única ontología es susceptible a cambios en las fuentes de información, los cuáles pueden afectar a la conceptualización del dominio representado en la ontología.

2. Utilizando Múltiples Ontologías

Al utilizar múltiples ontologías cada una de las fuentes de información es descrita por su propia ontología. En un principio la ontología fuente puede ser una combinación de varias ontologías, pero no se puede asumir que diferentes ontologías fuentes comparten el mismo vocabulario. En la figura 2.2 se describe esta arquitectura.

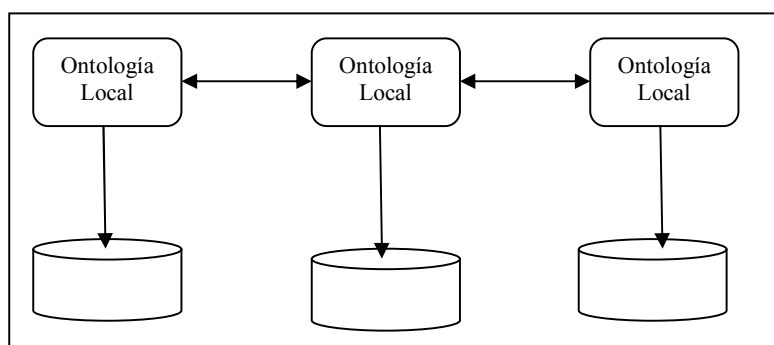


Figura 2.2. Integración utilizando múltiples ontologías.

En este enfoque cada ontología fuente se construye sin considerar otras fuentes de información o sus ontologías. Su arquitectura puede simplificar las tareas de integración y apoyar los cambios, por ejemplo, agregando o removiendo fuentes de información. Por otro lado, carece de un vocabulario común lo que dificulta poder comparar distintas ontologías fuentes. Para solucionar este problema es necesario definir una representación de mapping inter-ontología que identifica términos semánticamente correspondientes desde diferentes ontologías fuentes, donde los términos son semánticamente iguales o parecidos. Pero el mapping también tiene que considerar diferentes visiones sobre un dominio a través de diferentes agregaciones o granularidad de los conceptos de la ontología.

Para evitar una pérdida semántica, se tiene que permanecer dentro de la representación del lenguaje formal cuando se define un mapping entre distintas ontologías [13]. Una manera directa de permanecer dentro del formalismo es relacionar todas las ontologías usadas con una de nivel superior. Esto se puede hacer heredando los conceptos de la ontología de nivel superior. Este tipo de mapping puede utilizarse para resolver conflictos y ambigüedades. Además permite establecer conexiones entre los conceptos de diversas ontologías en términos de súper clases comunes, pero no establece una correspondencia directa. Porque puede conducir a problemas cuando se requiere una correspondencia exacta entre conceptos.

La correspondencia semántica intenta superar la ambigüedad que se presenta cuando se realiza un mapping indirecto de conceptos a través de un nivel superior, intentando identificar correspondencias semánticas bien fundamentadas entre los conceptos de las diferentes ontologías. Para poder procesar correspondencias entre campos de una base de datos se pueden utilizar etiquetas semánticas. Esto se puede lograr construyendo un modelo descriptivo lógico de los términos de las diferentes fuentes de información, demostrándose que el razonamiento se puede utilizar para establecer relaciones entre diferentes terminologías.

Para evitar mapping arbitrario entre los conceptos la correspondencia semántica maneja un vocabulario común para definir conceptos a través de diferentes ontologías.

3. Utilización de Ontologías Híbridas

Utilizar ontologías híbridas soluciona los problemas presentados en la utilización de ontologías únicas y múltiples. Este enfoque es similar al de múltiples ontologías ya que la semántica de cada fuente de información es descrita por su propia ontología. Pero para hacer comparables las ontologías locales cada una de ellas se construye desde un vocabulario global compartido. El vocabulario compartido contiene términos básicos (primitivos) de un dominio, el cual es combinado en la ontología local para describir

semánticas más complejas. En ocasiones el vocabulario compartido también es una ontología.

En el enfoque híbrido el punto interesante es cómo la ontología local es descrita. Una forma es asignar valores a los atributos dentro de un vector. Para este caso los términos derivan de una ontología de dominio global. Otra forma de describir la ontología local es ingresar cada concepto en etiquetas las cuales combinan los términos primitivos (raíz de una palabra) desde un vocabulario compartido. La combinación de operadores es similar a la de operadores conocidos desde una descripción lógica pero extendida, por ejemplo, un operador puede indicar que una palabra es una agregación de varias palabras separadas.

La ventaja de utilizar ontologías híbridas es que se pueden agregar nuevas fuentes de información con facilidad sin necesidad de hacer modificaciones. También apoya la adquisición y evolución de las ontologías.

La desventaja de este enfoque es que las ontologías existentes pueden no ser reutilizadas con facilidad, por lo tanto, se hace necesario que sean reconstruidas desde cero. La figura 2.3 describe la arquitectura de ontologías híbridas.

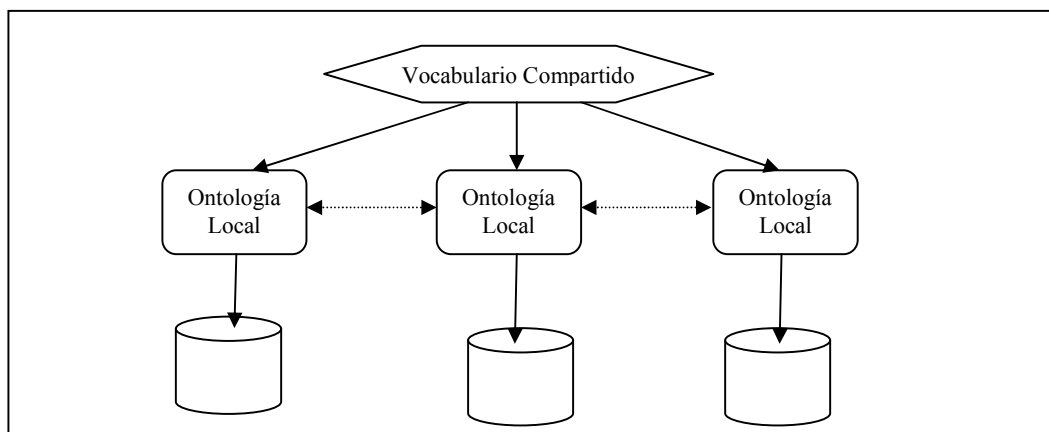


Figura 2.3. Integración utilizando ontologías híbridas.

En F. Wiesman [14] se establece un mecanismo para establecer la comunicación a través de mapping de ontologías dentro de un entorno de Sistemas Multiagentes (MAS. Multi Agents Systems). Donde los lenguajes de comunicación habilitan al agente para especificar la intención y el contenido del mensaje, así como el protocolo, el lenguaje y la ontología que ellos utilizarán.

Para explicar el mecanismo de mapping se supone el siguiente escenario en que dos agentes definen el concepto persona a través de conceptos, relaciones y atributos utilizando dos ontologías como se muestra en la figura 2.4 y 2.5. El agente 1 usa la ontología 1 y el

agente 2 utiliza la ontología 2. Ambas ontologías muestran algunas formas de heterogeneidad semántica

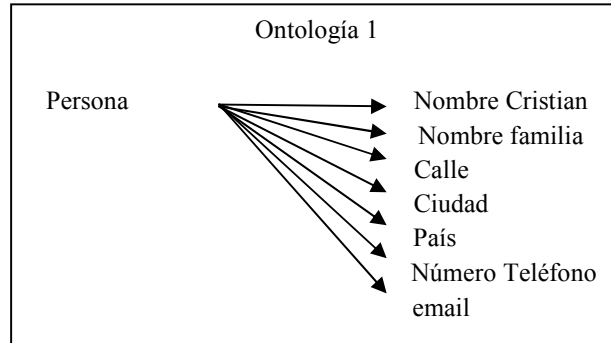


Figura 2.4. Representación ontología 1 usada por el agente 1.

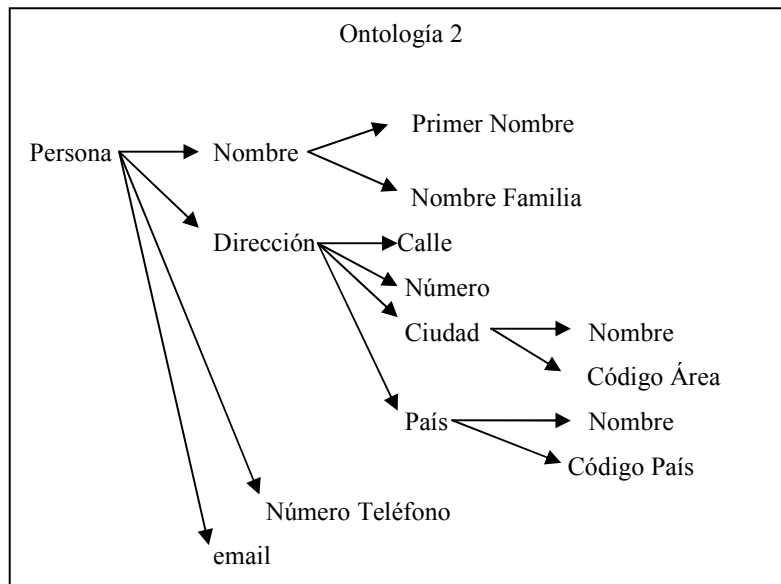


Figura 2.5. Representación ontología 2 usada por el agente 2.

En la ontología uno como se muestra en la figura 2.4 se describe el atributo “*calle*” como el número de la casa y el atributo “*número teléfono*” almacena el número, el código de área y el código de país. Mientras que en la ontología 2 mostrada en la figura 2.5 el atributo “*número teléfono*” almacena únicamente el número de la residencia, ya que el código de país y área son almacenados en el atributo “*ciudad*”.

Como se ve claramente cada una de las ontologías tiene distintas estructuras, la ontología 1 tiene una estructura plana mientras que la ontología 2 tiene una estructura jerárquica. Estos conflictos estructurales se resuelven con relativa facilidad ya que la

ontología 2 es más o menos una extensión de la ontología 1. Cuando dos ontologías tienen estructuras jerárquicas completamente distintas los conflictos estructurales son más serios.

Los conflictos de nombres entre las dos ontologías son problemas más graves. Ya que diferentes nombres se utilizan para representar los mismos tipos de datos, por ejemplo, primer nombre, y nombre Cristian. Por otra parte, un mismo nombre se utiliza para tipos de datos ligeramente diferentes, por ejemplo, calle. En la ontología 1 calle representa tanto el nombre de la calle y el número de la casa, mientras en la ontología 2 ésta sólo representa al nombre de la calle. Por lo tanto, para alcanzar la interoperabilidad se debe ser capaz de encontrar un mapping desde los conceptos de una ontología a los conceptos de otra ontología usando concatenación y división de operadores en el proceso de mapping.

Para establecer el mapping entre las dos ontologías no se pone ninguna restricción en la estructura de los conceptos. Un concepto puede ser definido como una agregación de atributos y sub-conceptos. Esta agregación puede incluso ser directa o indirecta, es decir, que contenga el concepto que está siendo definido o no.

Para explicar lo anterior se describe el siguiente escenario. El agente 1 desea conocer el teléfono y la dirección de una persona que se encuentra almacenada en la ontología definida por el agente 2. Por lo tanto, el agente 1 se contacta con el agente 2.

Para que se produzca la comunicación entre los agentes se debe establecer si éstos utilizan la misma ontología, o si ellos usan una ontología de la cual el otro agente sabe como mapearla a la suya. En el caso que utilicen ontologías diferentes y el mapping no se conozca, ellos tratarán de establecer uno.

Para establecer el mapping, primero se crea una representación plana, es decir, no estructurada de las instancias de conceptos. La representación plana se llama "*expresión*" y es utilizada en la comunicación entre los dos agentes. El intercambio de expresiones se utiliza para encontrar instancias de conceptos que representen la misma entidad en el entorno de los agentes.

El segundo paso es determinar las correspondencias entre las instancias de los conceptos en ambas ontologías, a través del intercambio de expresiones entre los dos agentes. La correspondencia entre las parejas de instancias forma la categorización de las entidades observadas.

El tercer paso es determinar la correspondencia entre los conceptos de las dos ontologías utilizando la categorización obteniéndose así las medidas de similitud. Por último establecer el mapping entre los conceptos correspondientes utilizando la categorización de las medidas de similitud.

El primer paso soluciona el problema del conflicto estructural, removiendo las estructuras de las ontologías, mientras que los dos últimos pasos manejan el problema del conflicto de nombre determinando el mapping más probable entre las ontologías, basado en las correspondencias de instancias de los conceptos.

Las *expresiones* representan toda la información relevante de una instancia, de una manera uniforme y contiene una identificación del concepto al cual pertenece la instancia representada para identificar el correspondiente concepto en la ontología. Ya que las dos ontologías pueden tener diferencias estructurales se consideran únicamente los valores de sus atributos.

Las jerarquías son aplanadas para representar cada uno de los atributos a través de etiquetas seguidas por el correspondiente valor del atributo. Las etiquetas deben ser únicas para cada trayectoria desde el concepto raíz al atributo. No es posible utilizar el nombre de un atributo en una etiqueta, porque puede suceder que el nombre del atributo se repita más de una vez en la definición de conceptos.

El valor de una etiqueta se representa por un string de caracteres. Esto garantiza que no se produzcan confusiones sobre su interpretación. Ya que un string pueda estar constituido por diferentes palabras, el valor de una etiqueta se puede interpretar como una lista de palabras separadas por un punto como muestra la figura 2.6.

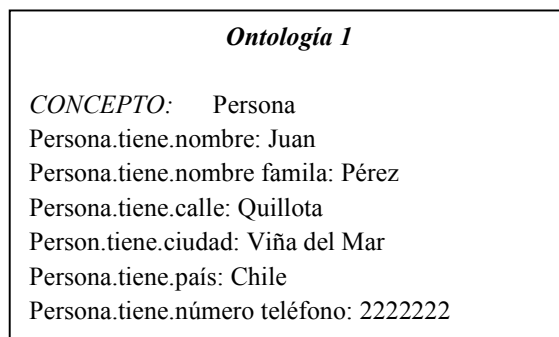


Figura 2.6. Representación de expresiones para la ontología 1.

Un aspecto importante es determinar cuanta información poner en una expresión. Puesto que el agente 1 quiere comunicarse con el agente 2 sobre el concepto persona debe decidir cuáles atributos y sub-conceptos del concepto persona incluirá en la expresión, es decir, qué parte de la ontología 1 será aplanada y eso dependerá de las necesidades de información del agente 1.

Una vez que el agente 2 recibe una expresión u^1 desde el agente 1, éste debe aplanar la parte de su ontología que generará una mejor correspondencia, obteniéndose la expresión

u^2 . Para ello utiliza *técnicas de Information Retrieval* para los datos no estructurados, con el objetivo de determinar la mejor correspondencia con su expresión u^2 . Es decir, el agente 2 busca para instancias de un concepto cuáles sub-conceptos proporcionan información relevante. La correspondencia de pares de instancias que se identifiquen formarán las entidades observadas.

La generación de las entidades observadas la establece el agente 2 expresión u^1 enviada por el agente 1, donde busca qué instancias similares de la ontología 2 serán representadas por la expresión u^2 . Para este proceso de búsqueda se utiliza un análisis probabilístico que indica la probabilidad que dos instancias representadas por sus expresiones representen la misma entidad en el mundo, dada la correspondencia de palabras de ambas expresiones. La medida de similitud es obtenida estableciendo probabilidades condicionales que evalúan que las dos expresiones representen la misma entidad en el mundo dada la correspondencia de palabras.

Una vez que se tienen los pares de entidades observadas, es decir, la medida de similitud, se establece la correspondencia de conceptos en las dos ontologías usando este valor. El agente 2 decidirá si dos expresiones representan la misma entidad en el mundo utilizando un valor umbral θ con ello asegurará que para cada uno de los conceptos en la ontología 1 halla a lo más un concepto en la ontología 2, así el agente 2 considerará que un concepto dentro de la ontología 2 es idéntico a uno de la ontología 1 si $\theta > 1$. Para así finalmente realizar el mapping entre las dos ontologías.

El ejemplo que se describió da una primera idea de las técnicas que podrían ser útiles para poder comparar competencias, es decir, hacer comparaciones semánticas, que indiquen el grado de similitud entre competencias de distintas carreras a nivel técnico, que es lo buscado por este proyecto.

2.4 Knowledge Sharing

Dentro de M. Uschold [20], se hace un análisis comparativo de tres formas de conocimiento compartido, servicios compartidos vía traducción punto a punto, formatos neutrales de intercambio y autorización neutral.

En M. Uschold [21], se analiza la necesidad común de compartir el significado de los términos en un dominio dado. Como resultado se obtiene una ontología que puede ser aplicada en una amplia variedad de contextos para varios propósitos.

2.4.1 Servicios Compartidos Vía Traducción Punto a Punto

En este enfoque un sistema pone sus servicios a disposición de otro. Para poder hacer uso de esos servicios se requiere de una traducción, porque los conceptos y términos usados por cada uno de los sistemas no son iguales. Primero cada sistema usa diferentes representaciones del lenguaje, por lo tanto el conocimiento debe ser traducido para que sus propiedades dinámicas de comportamiento sean conservadas. Segundo, aunque los lenguajes fueran iguales, el conocimiento en cada sistema puede basarse en diferentes dominios ontológicos, por consiguiente se requiere que haya una traducción en términos de dominio durante la transferencia del conocimiento.

Para evitar compartir el comportamiento del conocimiento completo, se utilizará un modelo de delegación vía servicios compartidos, para la solución de problemas. Es decir, un sistema que requiera una respuesta a una pregunta de diseño en particular no solicitará y no importará el conocimiento para contestar a esa pregunta, sino que por el contrario delegará el problema a otro sistema capaz de contestarle. Esto simplifica el problema de la traducción, ya que sólo las preguntas y las respuestas necesitan ser traducidas entre los diversos sistemas, no la base de conocimiento completa. Este modelo se basa en la metáfora de una comunidad de expertos, colaborando para solucionar juntos los problemas sin intentar enseñar cada uno su experiencia.

Como modelo de pruebas se usaron dos sistemas de la compañía aérea Boeing. Uno de ellos fue un sistema experto para la selección de materiales ESDS y otro para el diseño generativo basado en el sistema Génesis. Este sistema en ocasiones requiere consultar por materiales particulares para la construcción de algún elemento, por ejemplo, la construcción de un tubo. Para ello se inicia una sesión de preguntas y respuestas en tiempo real, entre los dos sistemas, hasta que finalmente el sistema experto da una recomendación de material.

Como ya se ha mencionado las diferencias entre los dominios de cada ontología son una barrera para poder compartir conocimiento. Por lo tanto, para hacer posible el intercambio de conocimiento, primero hay que hacer un cuidadoso análisis a cada uno de los sistemas e identificar los términos y conceptos usados por ellos. Segundo hay que crear un conjunto de reglas de mapping que indiquen como mapear los términos de una ontología a otra, para ello se debe construir un mediador. El tercer paso es construir el mediador.

La figura 2.7 muestra el modelo simplificado de servicios compartidos vía traducción punto a punto. La figura 2.8 muestra el proceso completo de servicios compartidos vía traducción punto a punto.

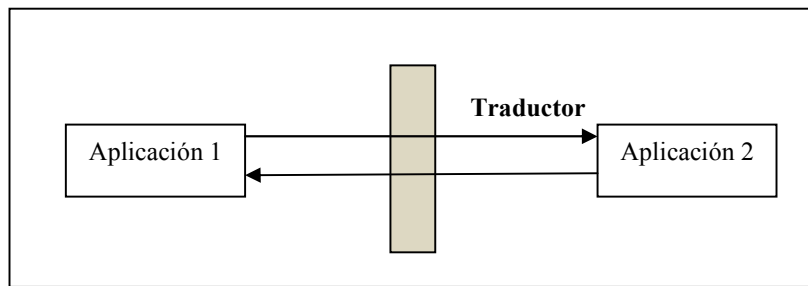


Figura 2.7. Servicio compartido vía traducción punto a punto modelo simplificado.

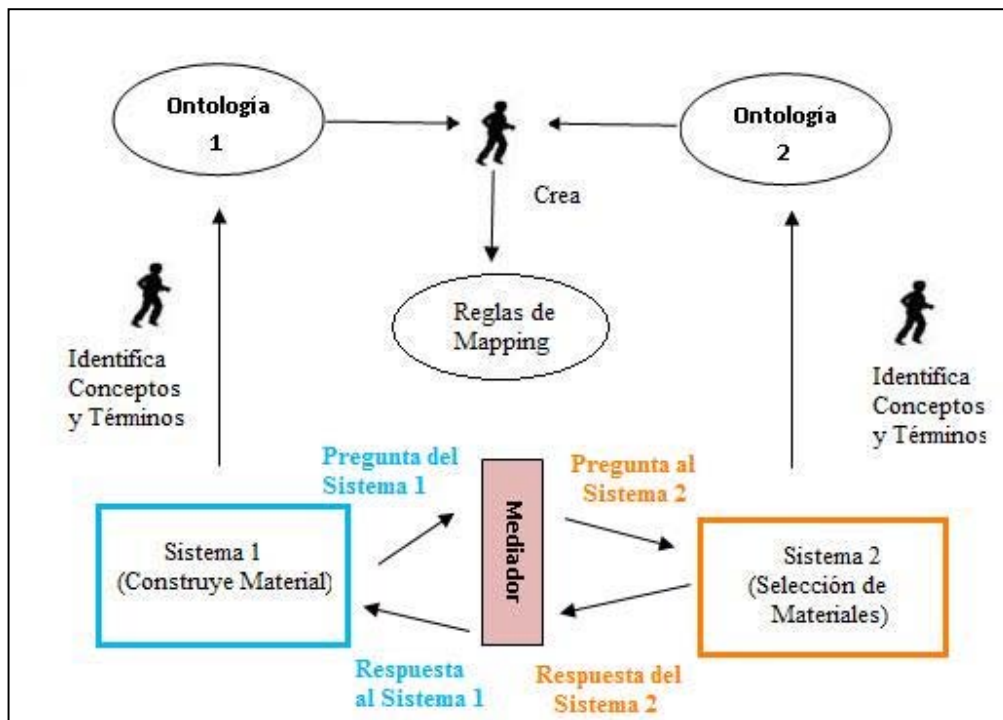


Figura 2.8. Servicio compartido vía traducción punto a punto proceso completo.

2.4.2 Formato de Intercambio Neutral

Esta forma de intercambio de conocimiento es más útil cuando se quieren involucrar más de dos sistemas. Para su diseño se pueden seguir varias estrategias una de ellas es hacerlo lo suficientemente expresivo para no perder nada en el proceso de intercambio entre la aplicación objetivo y el formato neutral, por ejemplo que su expresividad cubra la expresividad de todos los lenguajes objetivos individuales. Por ejemplo si la aplicación J necesita información que tiene la aplicación K , la información debe ser traducida desde la aplicación J al formato neutral, y desde el formato neutral a la aplicación K .

Otra estrategia es construir traductores de dos vías entre el formato neutral y cada uno de los formatos de aplicación objetivos.

Este modelo de conocimiento compartido permite que las aplicaciones utilicen la información de otras aplicaciones, con varias ventajas potenciales. Primero, no hay necesidad de construir aplicaciones para aprender un nuevo lenguaje para que se pueda autorizar el intercambio, puesto que la autorización se realiza en el formato de la aplicación original. En segundo lugar, los diversos sistemas se pueden mantener independientemente, por lo menos en teoría, ya que si hubiera que cambiar el lenguaje de la aplicación se modificaría únicamente el traductor cambiando su propio formato al del formato neutral.

El formato de intercambio neutral va a ser posible dependiendo de la naturaleza de la información que se intercambie, como para aplicaciones de datos que no tienen propiedades de comportamiento. Para el caso de intercambio de datos con propiedades de comportamiento este proceso no ha madurado lo suficiente. La figura 2.9 muestra el esquema del formato de intercambio neutral.

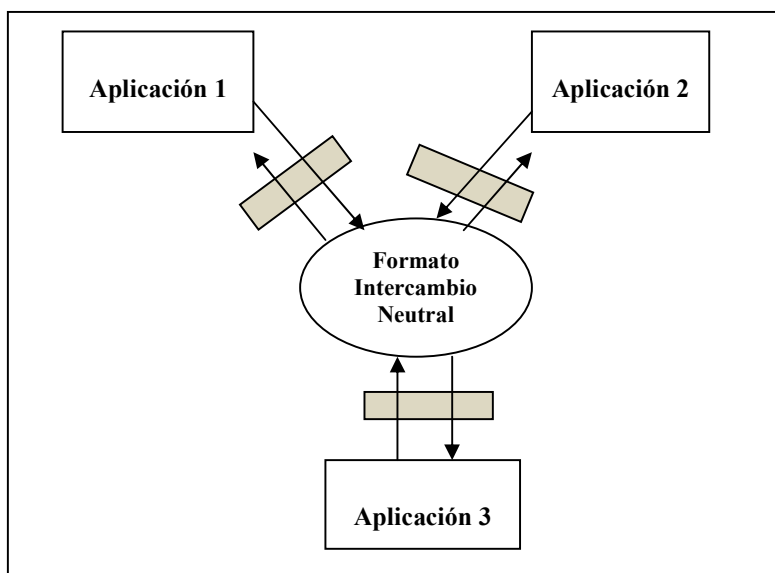


Figura 2.9. Formato de Intercambio Neutral.

2.4.3 Autorización Neutral

Es similar al formato de intercambio neutral, sin embargo, en este enfoque el conocimiento se autoriza en el formato neutral, y la traducción sólo se realiza de una única manera, dentro del formato objetivo. Es decir, se autoriza la utilización de un único lenguaje, en lugar de utilizar múltiples lenguajes objetivos. Nada se traduce en el formato neutral. Las condiciones de diseño son similares a las del formato de intercambio neutral.

Hay varios beneficios en su utilización. Primero, la autorización es en un sólo formato y la traducción es dentro de múltiples lenguajes objetivos, por lo tanto, se reduce la dependencia de los formatos particulares. Además, se necesita mantener únicamente una

versión del conocimiento que contribuye a una mantención potencialmente más barata y eficaz.

Para analizar en detalle este tipo de intercambio se realizaron una serie de experimentos usando una herramienta llamada “*Specware*”, que se utiliza para la especificación y desarrollo formal de software, en el cual, el primero es especificado utilizando un lenguaje de alto nivel, llamado “*SLANG*”, y entonces esta especificación es interactivamente traducida a un código ejecutable vía un proceso formal de refinamiento. Los resultados de los experimentos demostraron que este intercambio es posible para algunas partes muy simples de la aplicación, que incluyan atributos simples y sub-partes, y desarrollan un lenguaje neutral para describirlas. Éstas se pueden traducir dentro de dos lenguajes “*KBE (Knowledge-Based Engineering)*”, cargándolas y ejecutándolas con éxito en los sistemas de software objetivos. Sin embargo, este éxito inicial fue limitado al usar las características que eran en esencia iguales en ambos lenguajes.

Un traductor diseñado en formato neutral para la autorización de ontologías es Ontolingua.

La aplicación de este proceso sufre de las dificultades de la traducción y el costo del diseño debido al lenguaje, por lo tanto no es todavía una opción posible, pero hay casos en los que sí se puede aplicar, y es cuando el lenguaje objetivo está restringido. La figura 2.10 muestra el modelo de autorización neutral.

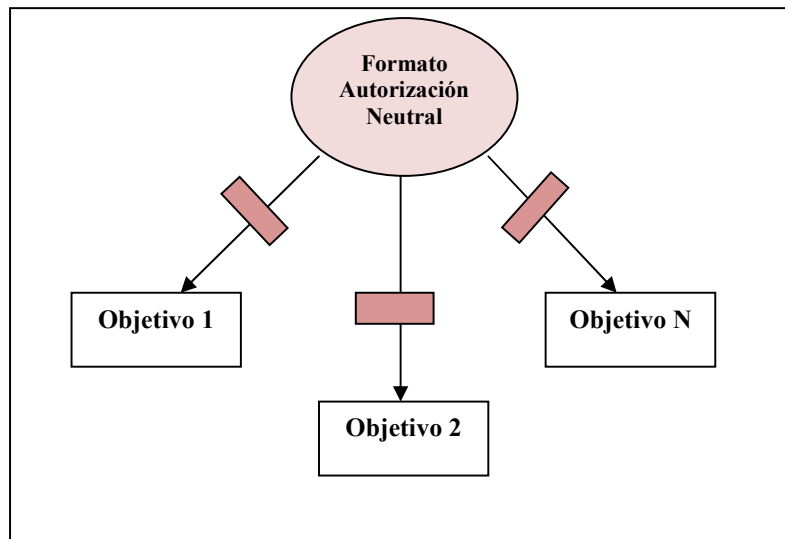


Figura 2.10. Modelo de Autorización Neutral.

2.4.4 Ontologías para Compartir Conocimiento

En el enfoque que se analiza en M. Uschold [21], se realiza un análisis de la aplicación de ontologías para la búsqueda y recuperación de información entre distintas fuentes. Debido a la necesidad imperante de compartir el significado de algunos términos dentro de un dominio dado. Alcanzar un entendimiento se logrará en la medida que se agregue conceptualización al dominio y sólo entonces se hará explícito a algunos lenguajes. El resultado de esto dará una ontología que se pueda aplicar en una amplia variedad de contextos para varios propósitos.

Una ontología puede tomar muchas formas, pero necesariamente debe incluir un vocabulario de términos y alguna especificación de su significado. Incluyendo definiciones e indicando como se relacionan los conceptos que en conjunto darán estructura al dominio y limitarán las posibles interpretaciones de los términos.

En este trabajo se presenta un framework que propone una nomenclatura común, para lograr la comprensión y clasificación de las aplicaciones de las ontologías, identificando las principales categorías de éstas, que son: autorización neutral como se describió anteriormente, acceso común a la información e indexación para la búsqueda. Para cada una de las categorías se identifican uno o más escenarios de aplicación. Para el framework propuesto se distinguen las siguientes dimensiones claves:

- **Propósitos y beneficios:** Se destaca el objetivo de utilizar ontologías para permitir la comunicación entre los humanos y los sistemas computacionales, se enfatizan las tres áreas principales de aplicación de las ontologías que son: primero para la comunicación entre humanos donde una ontología informal es suficiente. Segundo la interoperabilidad entre sistemas computacionales, donde se requiere una ontología como un formato de intercambio. Tercero mejorar los procesos o calidad de los sistemas de software.

Las ontologías son la base para la codificación formal de las entidades, atributos, procesos y relaciones más importantes en el dominio de interés. Esta representación puede ser un componente reutilizable o compartido en un sistema de software.

Una ontología puede utilizarse como un metadato que actúe como índice en un repositorio de información.

Una representación formal hace posible automatizar el control de consistencia lo que da como resultado un software.

Una ontología puede asistir al proceso de identificación de requerimientos y definir una especificación para sistemas de tecnologías de información.

Usar una ontología en el desarrollo de un sistema, o como parte de una aplicación final, puede facilitar la mantención de muchas maneras. Sistemas que se construyen usando una ontología explícita sirven para mejorar la documentación del software reduciendo los costos de mantención. La mantención también es un beneficio importante si una ontología se utiliza como lenguaje de intercambio neutral con múltiples lenguajes objetivos, ya que la mantención se hace sólo en un lugar.

La velocidad y fiabilidad se puede incrementar utilizando una ontología existente como punto de partida y base para guiar la adquisición de conocimiento cuando se construyen sistemas basados en conocimiento.

- **Rol de la Ontología:** Dentro del escenario de aplicación la información puede jugar tres roles diferentes, ser un dato operacional, una ontología o un lenguaje de representación de ontologías dependiendo del punto de vista. Además el escenario de aplicación de la ontología varía según como se utilice ésta, ya que para un escenario dado más de una ontología puede estar involucrada.
- **Actores:** Cada escenario involucra un conjunto de actores y cada uno de éstos representa un rol que una persona o aplicación puede jugar, a la vez, éstos pueden jugar más de un rol en un escenario. Los actores pueden ser: el autor de la ontología (AO), el autor de los datos operacionales (AD), el desarrollador de aplicaciones (DA), el usuario de la aplicación (UA) y el trabajador de conocimiento (TC).
- **Tecnologías de Apoyo:** Hay una gran variedad de tecnologías que apoyan la aplicación de ontologías, cómo los lenguajes de representación de ontologías (UML, Express, Ontolingua, XML), lenguajes de intercambio de conocimientos (KIF, PIF, CDIF), traductores de herramientas (Ontolingua translator, CDIF-tools, Step tools, etc.) y objetos distribuidos (CORBA, COM, entre otros).

La primera categoría analizada es la autorización neutral donde un sistema de información, que puede ser una ontología o un dato operacional, es autorizado en un lenguaje único y convertido a una forma diferente, esta traducción puede ser, directa lenguaje a lenguaje, manual o automatizada, en cualquier caso el proceso se hace explotando la sintaxis o la semántica de los conceptos representados, para que la ontología o el dato pueda ser usado por múltiples sistemas. El beneficio de esto es la reutilización y portabilidad del conocimiento a través de las aplicaciones, mejorando la mantención y la retención del conocimiento.

Un primer escenario de aplicación dentro de la categoría mencionada es la autorización de ontologías, su objetivo es disminuir el costo de la reutilización e

incrementar la mantención del conocimiento. Para cumplir estos objetivos, los actores desarrollan una ontología, la cual se pueda convertir dentro de muchos sistemas operacionales. Las tecnologías de apoyo incluyen traductores unidireccionales de la ontología y desarrollo de software de procesos (KADS). Los actores principales del escenario son los usuarios de la ontología y los autores de ésta. La figura 2.11 muestra la autorización de ontologías.

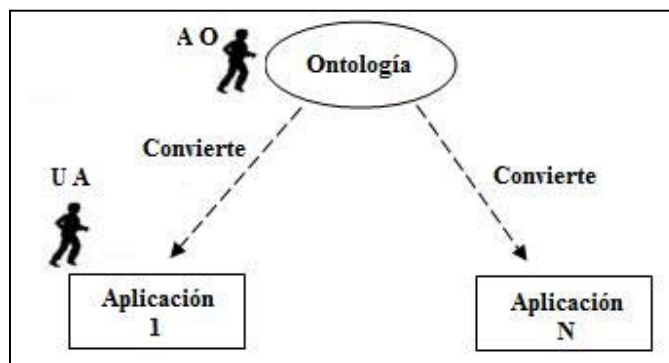


Figura 2.11. Autorización Neutral de ontologías.

El segundo escenario es la autorización de datos operacionales, en este caso la estructura es similar a la presentada en la autorización neutral de ontologías, pero el enfoque está en la traducción y autorización de los datos operacionales más que en la ontología. La diferencia principal es el rol que juega la ontología en el escenario y quiénes son los actores principales.

La autorización neutral de datos operacionales mejora la mantención y transportabilidad de estos datos. Para lograrlo el autor de la ontología, que es un actor secundario para este escenario, crea una ontología la cual define el formato neutral utilizado por el actor primario, que es el autor de los datos operacionales. Las herramientas traducen estos datos a un sistema objetivo, usando un traductor unidireccional. Entonces los usuarios de la aplicación interactúan con el sistema para ejecutar un análisis o consulta de los datos operacionales.

Una única ontología puede convertirse y utilizarse en muchas aplicaciones diferentes, por lo tanto, es una manera importante de alcanzar la reutilización del conocimiento. Si varios sistemas se basan en la misma ontología, entonces ésta facilita la interoperación entre los sistemas si es requerido. Sin embargo, no implica el compartimiento o intercambio de conocimiento entre los sistemas. La figura 2.12 muestra la autorización de datos operacionales.

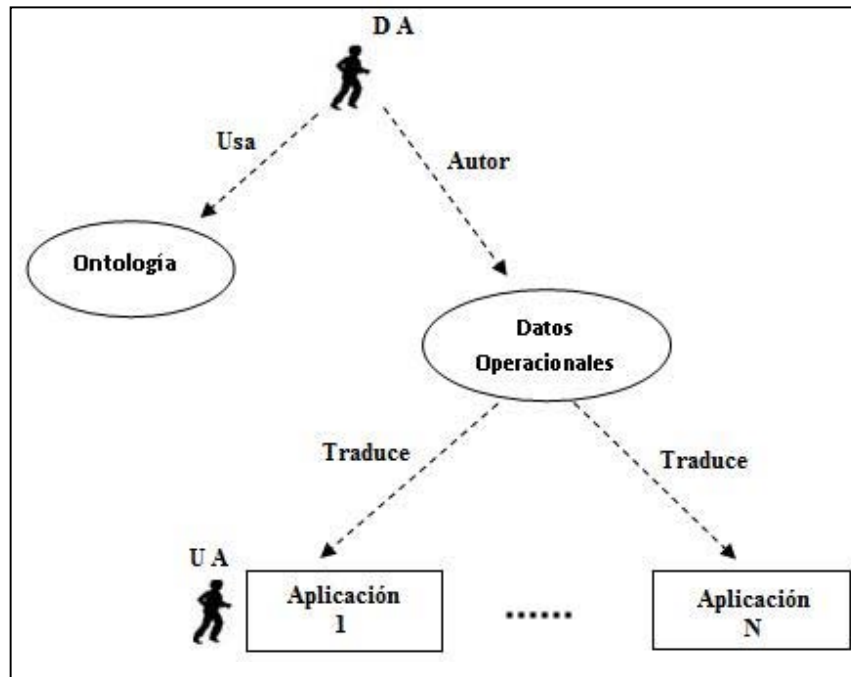


Figura 2.12. Autorización de Datos Operacionales.

Una segunda categoría es la correspondiente al acceso común a la información, donde la idea es utilizar ontologías para habilitar múltiples sistemas objetivos para que tengan acceso a fuentes heterogéneas de información. El beneficio de este enfoque es la reutilización y la interoperabilidad del conocimiento.

En los escenarios que se describirán para esta categoría los interesados en la información pueden ser humanos o sistemas. En segundo lugar la información puede ser una ontología o datos operacionales que pueden o no ser computacionales.

El primer escenario de aplicación para esta categoría es la comunicación entre humanos, donde el mayor beneficio de desarrollar una ontología es promover el entendimiento común entre los trabajadores del conocimiento (TC). Para lograrlo los diseñadores o autores de la ontología desarrollan una ontología compartida, a la cual los otros trabajadores del conocimiento pueden hacer referencia durante su trabajo.

Las habilidades no computacionales, tales como la clasificación de librerías son valoradas en la construcción de estas ontologías, que comúnmente toman la forma de glosarios. Las tecnologías utilizadas para el desarrollo incluyen editores de ontologías y browsers. Los actores principales son los diseñadores o autores de éstas y los trabajadores del conocimiento. La figura 2.13 muestra la comunicación entre humanos.

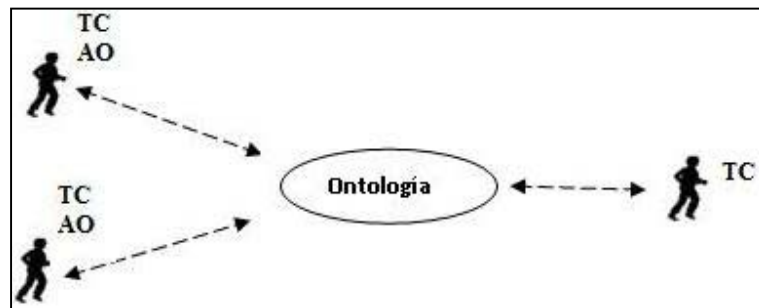


Figura 2.13. Comunicación entre Humanos.

El segundo escenario de aplicación dentro de esta categoría es el acceso de datos vía ontología compartida. El objetivo de este escenario es reducir el costo de múltiples aplicaciones teniendo acceso común a los datos, facilitando la interoperabilidad. Esto se puede lograr si los desarrolladores acuerdan una ontología compartida, la cual define un lenguaje común para intercambiar o compartir datos operacionales. Las tecnologías utilizadas para el desarrollo de la ontología incluyen traductores y gestores de análisis. Los actores principales son los autores o desarrolladores de la ontología y los desarrolladores de las aplicaciones.

Para su implementación un diseñador o autor de ontologías creará una donde diferentes desarrolladores de aplicaciones acuerdan su uso. Éstos definen un formato de intercambio, para el cual se requieren dos maneras de traducción entre el formato de intercambio y el formato de aplicación. Cada traductor define una interfaz que puede ser usada para leer y escribir datos. Frecuentemente los traductores se crean manualmente o se utilizan lectores y escritores explícitos que son creados automáticamente usando gestores de análisis. La interoperabilidad entre las múltiples aplicaciones se hace posible al permitir el acceso al mismo tipo de información. La figura 2.14 muestra el acceso de datos vía ontología compartida.

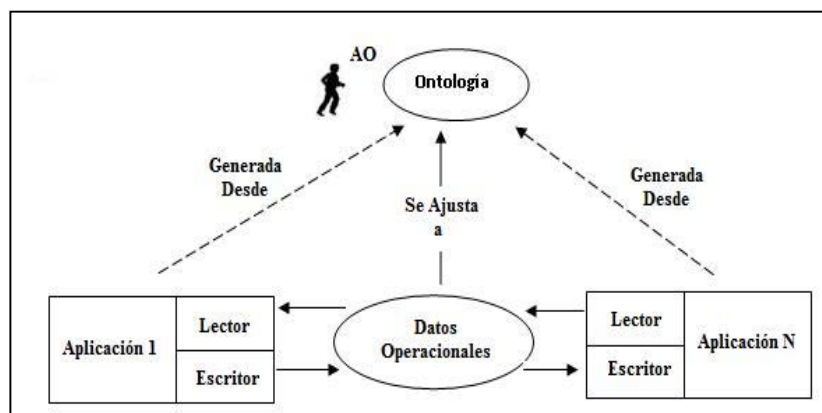


Figura 2.14. Acceso de Datos Vía Ontología Compartida.

El tercer escenario dentro de la categoría de acceso común a la información es el acceso de datos vía mapping de ontologías. En este escenario no hay una ontología compartida explícita, en su lugar, se usan reglas de mapping para determinar qué términos en una ontología tiene un significado en la otra. Para ello se utiliza un mediador que utiliza estas reglas en *tiempo real* para que todas las aplicaciones puedan acceder a los datos entre sí.

La ventaja de este enfoque es que no requiere un desarrollador de aplicaciones explícito para acordar una ontología compartida. Las tecnologías que apoyan el desarrollo de este método utilizan, generadores de análisis y mediadores. Los actores principales son los desarrolladores de las ontologías y los desarrolladores de las aplicaciones.

Cada una de las aplicaciones que desean intercambiar datos tiene su propia ontología local. Los desarrolladores cooperan creando un mapping compartido que relaciona términos en diferentes ontologías. Este mapping se utiliza para generar un mediador, que mapea los datos operacionales expresados en la terminología de una ontología con los datos operacionales expresados en la otra ontología. La figura 2.15 muestra el acceso a los datos vía mapping de ontologías.

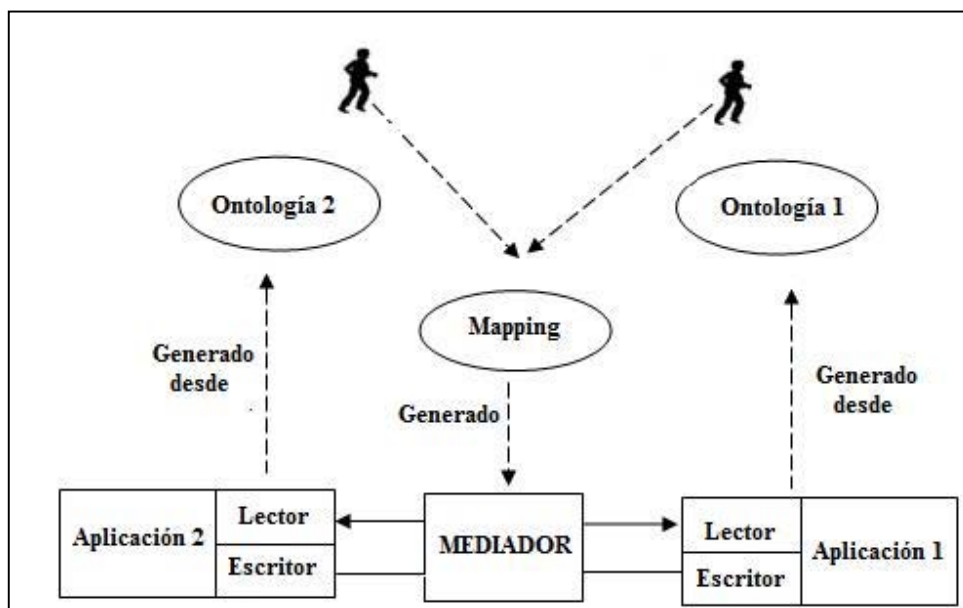


Figura 2.15. Acceso de Datos Vía Mapping de Ontología.

El último escenario dentro de la categoría de acceso común a la información es el de servicios compartidos. Este enfoque busca en la compartición de servicios la neutralidad, por ejemplo, en el lenguaje, las máquinas, los sistemas operativos y la ubicación. Para ello los desarrolladores acuerdan una ontología compartida, la cual define las interfaces para

múltiples lenguajes objetivos. Las tecnologías que apoyan esta visión utilizan generadores de interfaces. Los actores principales son los autores de la ontología y los desarrolladores de aplicaciones.

Para implementar este escenario un autor o desarrollador de ontologías crea una ontología que diferentes desarrolladores de aplicaciones acuerdan usar. Éstas se utilizan para generar la interfaz de aplicación y las definiciones que cada aplicación utiliza para leer y escribir datos.

Esta perspectiva de trabajo es muy similar al acceso de datos vía ontología compartida. El éxito de la aplicación depende principalmente del acuerdo sobre una ontología que contenga suficiente riqueza semántica para satisfacer los requerimientos del cliente y el servidor. La figura 2.16 muestra este escenario de aplicación.

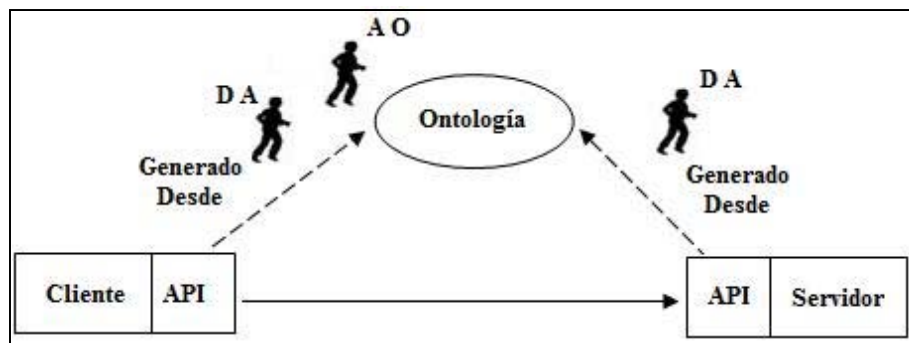


Figura 2.16. Servicios Compartidos.

La última categoría en análisis es la indexación y su escenario de aplicación es la búsqueda basada en conceptos. Este escenario se basa en la ubicación de documentos en algunos repositorios. Para lograrlo los trabajadores del conocimiento utilizan una ontología como motor de búsqueda que se ejecuta como un índice dentro del repositorio.

Las tecnologías que apoyan este enfoque incluyen navegadores de ontologías y motores de búsqueda. Los actores principales son los autores de ontologías y los trabajadores del conocimiento.

En este escenario un desarrollador de ontologías creará una que será usada por los diferentes trabajadores del conocimiento para identificar los conceptos que a ellos les interesan. Un motor de búsqueda usa estos conceptos para ubicar los documentos relevantes dentro de un repositorio. La figura 2.17 ejemplifica la aplicación de este escenario.

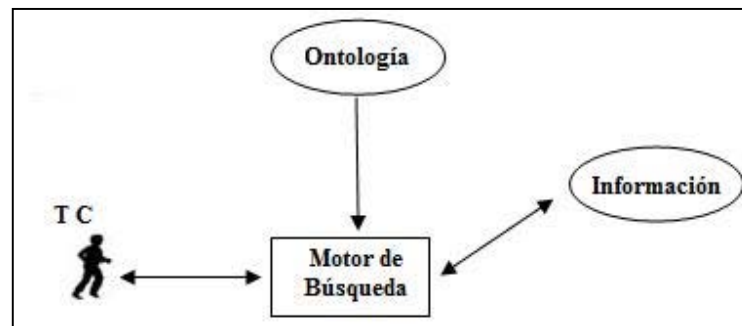


Figura 2.17. Búsqueda Basada en Conceptos.

2.4.5 Mediadores

En H. Wache [22], se analiza como las reglas basadas en mediadores se pueden extender para resolver problemas de heterogeneidad semántica.

Los mediadores se utilizan para la integración de fuentes semánticas heterogéneas. Un mediador proporciona una integración flexible de varios sistemas de informáticos tales como, sistemas administradores de base de datos, sistemas de posicionamiento geográficos o incluso para la W.W.W. Ya que éstos combinan, integran y abstraen la información proporcionada por las fuentes. Normalmente las fuentes de información son encapsuladas por envolturas, que ocultan el modelo interno de la estructura de datos y transforman el contenido a un modelo de estructura de datos uniforme.

Las reglas de transformación basadas en mediadores, describen cómo la información de las fuentes pueden ser mapeadas a una vista integrada proporcionada por el mediador. Los mediadores utilizan estas reglas para dividir las consultas, que se formulan basándose en la vista integrada, en sub-consultas para cada fuente de información y combina los resultados devueltos por la fuente de acuerdo a la consulta.

Las reglas basadas en mediadores se diseñan principalmente para resolver los problemas de heterogeneidad estructural, es decir, que diferentes sistemas almacenan sus datos en diferentes estructuras. Para resolver este problema se debe considerar el nivel semántico y una forma de hacerlo es a través del contexto. Un contexto contiene metadatos relacionados a su significado, propiedades (como su fuente, calidad y precisión) y organización. Un valor tiene que ser considerado en su contexto y puede ser transformado a otro, lo que se conoce como transformación del contexto.

En el trabajo de H. Wache [22], se proponen reglas basadas en la representación formal para la transformación de contexto y para la integración. Para la transformación de contexto se utilizan las reglas de contexto.

Las reglas de contexto aplican la transformación de contexto en una representación formal declarativa y expresiva. Las reglas de contexto y de integración componen el conjunto de reglas de transformación las cuales configuran las reglas basadas en mediadores.

Las reglas de transformación guían el mapping de la fuente de información a una vista integrada proporcionada por el mediador. Las reglas de transformación se construyen en base a *template*. Los *template* son objetos genéricos, centrados en la descripción de la información. Tienen un nombre, un contexto que trata la semántica del concepto, un tipo que determina el tipo de dato, un valor que contiene el valor (número, string, lista de atributos) y por último un identificador que indica a qué fuente de información pertenece el *template*. Un atributo consiste de un nombre y un *template* al atributo de referencia. Un *template* se representa de la siguiente manera:

$$T = \langle nombre, contexto, tipo, valor \rangle @ fuente \quad (2.1)$$

El tipo y el valor describen la estructura de la información, mientras que el contexto captura la semántica de la información. El contexto es diseñado para tomar conflictos semánticos obvios y guiar a la solución de los problemas de heterogeneidad semántica. Su representación se basa en un conjunto de meta atributos, que constan de un nombre y un valor que pueden derivarse de una ontología común, asegurando una semántica global sobre toda la fuente de información. Un contexto se escribe de la siguiente manera:

$$C = \langle (C_1, V_1), (C_2, V_2), \dots, (C_n, V_n) \rangle \quad (2.2)$$

Un *template* con atributos puede representar tablas (relaciones) en un modelo relacional de estructura de datos. Los atributos del *template* son los atributos de la relación. Los valores de los atributos del *template* son *templates* que encapsulan tipos de datos básicos de la relación de atributos.

Se pueden distinguir dos clases de reglas de transformación, las reglas de combinación y las reglas de reemplazamiento. Cada regla de transformación consta de una cabeza y un cuerpo. El cuerpo consiste de un conjunto de *template* que son conectados a través de conjunciones, mientras que la cabeza se construye sobre un *template*.

Una regla de combinación es diseñada para la combinación e integración de información, principalmente para la solución de problemas de heterogeneidad estructural, pero sólo pueden ser aplicadas a *template* de alto nivel. La información es ubicada de acuerdo al *template* en el cuerpo de la regla. Entonces la información es construida con

respecto a la cabeza de la regla. Es decir, que la información se propaga a través de variables en el *template*.

Los ítems de información son limitados a las variables en el cuerpo del *template*. Entonces éstos son propagados a la cabeza y cambiados de acuerdo a la estructura de la cabeza del *template*. Una regla de combinación se define de la siguiente manera:

$$H \leftarrow B_1, \dots, B_n, \emptyset_1, \dots, \emptyset_m \quad (2.3)$$

H y B_1, \dots, B_n son *templates*; $\emptyset_1, \dots, \emptyset_m$ son expresiones. Las expresiones son formuladas sobre las variables en el *template* y restringen las posibles sustituciones de éstas.

Las reglas de reemplazamiento son necesarias para que se pueda realizar el intercambio de información, ya que pueden transformar ésta en otra estructura. A la vez, las reglas también pueden ser aplicadas a una información que esté jerarquizada en una estructura de alto nivel, por ejemplo un atributo. Las reglas de reemplazo sustituyen la información en el atributo con la información transformada. Se representan de la siguiente manera:

$$H \leftarrow \bar{B} \& B_1, \dots, B_n, \emptyset_1, \dots, \emptyset_m \quad (2.4)$$

En las reglas de reemplazo un *template* \bar{B} en el cuerpo de *templates*, son necesarios para que sean firmados. Un *template* firmado reemplaza al *template* cabeza H de la regla. Los otros *template* en el cuerpo B_1, \dots, B_n y las expresiones $\emptyset_1, \dots, \emptyset_m$ se requieren para apoyar o restringir el reemplazo.

Los párrafos anteriores describieron las dos clases de reglas de transformación que representan las tareas de integración y la transformación del contexto. En el nivel de la aplicación el conjunto de reglas de transformación se divide en reglas de integración y reglas de contexto que describen qué reglas serán usadas para una tarea.

Para la integración, las fuentes de información son mapeadas a la vista integrada proporcionada por el mediador. Para cada fuente de información existe una regla de combinación que mapea una fuente a una vista integrada. Si una fuente no proporciona la suficiente información requerida por la vista integrada, entonces esa fuente se integra con otra que proporciona la información faltante.

Durante la transformación del contexto, la información también es transformada. La formulación de una regla de contexto no depende de la ubicación de la información que

será convertida. Dado que las reglas de reemplazo permiten intercambiar información incluso en estructuras de *templates* jerarquizados, son el formalismo de representación adecuado para las reglas de contexto.

Como se mencionó en los párrafos anteriores los mediadores utilizan las reglas de transformación para dividir las consultas en sub-consultas, que son formuladas con respecto a la vista integrada. Una consulta consiste de un conjunto de templates derivados de la vista integrada. Las variables en el template de la consulta indican qué información es relevante para la respuesta, además el template de la consulta define el contexto en el cual se debe dar la respuesta. Un template de consulta se representa de la siguiente manera:

$$Q = \{ T_1, T_2, \dots, T_n, \emptyset_1, \emptyset_2, \dots, \emptyset_m \} \quad (2.5)$$

Las expresiones también se deben considerar con respecto al contexto de la consulta. Por lo tanto, las expresiones antes de ser enviadas a la fuente de información tienen que ser transformadas al contexto de la fuente, de lo contrario se llegará a una respuesta incorrecta.

La utilización de mediadores para la integración de fuentes de información puede servir de gran utilidad, porque ayuda a resolver el problema de heterogeneidad semántica antes mencionado en este capítulo. Por lo tanto podría pensarse como una posible solución, pero debió ser descartada porque los mediadores son diseñados específicamente para juntar información de varios repositorios distintos, como forma de dar respuesta a una inquietud del usuario, y lo que este proyecto busca es poder comparar dos certificados distintos de manera de poder determinar si un alumno cumple con los conocimientos adecuados para poder completar parte de sus estudios en otro país.

2.5 Information Retrieval

Según N. Belkin [15], los sistemas de Information Retrieval (I.R) cumplen con la función de guiar al usuario a aquellos documentos que lo ayudan a satisfacer sus necesidades de información dentro de una larga base de datos de documentos. Una definición un poco más general indica que el objetivo de estos sistemas es dar al usuario información desde fuentes de conocimientos que lo ayudarán en problemas de administración.

Los documentos almacenados generalmente están en formato de texto. A diferencia de las técnicas convencionales de base de datos, las técnicas de I.R son útiles cuando la semántica de los objetos que van a ser recuperados es poco clara y las relaciones entre los atributos de los objetos y las necesidades de información de los usuarios son complejas o poco comprendidas. Para Lewis [18], el proceso I.R contiene 4 fases principales:

- 1. Indexación:** Los documentos originales deben ser convertidos por algunos métodos de representación de texto dentro de un formato, algunas veces llamado documento representativo, los cuales son utilizados por los sistemas de software de I.R. La mayoría de los sistemas de I.R representan textos como vectores binarios o numéricos derivados de las palabras del documento o asignados por indexadores humanos.
- 2. Formulación de consultas:** Los usuarios deben expresar sus necesidades de información a través de consultas interpretadas por los software de I.R. La consulta algunas veces es ingresada por el usuario de una manera muy similar a la usada por el sistema, como una expresión booleana de las palabras claves. En otros casos, la conexión puede ser menos directa, por ejemplo consultas ingresadas en lenguaje original o conjunto de documentos y el software de I.R selecciona las palabras importantes de las ingresadas por el usuario para usarlas en un vector estadístico de las palabras que más aparecen en un documento.
- 3. Comparación:** El sistema debe comparar las consultas del usuario con los documentos almacenados y tomar una decisión sobre qué documentos recuperar y en qué orden. Las consultas proporcionan la información necesaria para diferenciar entre documentos relevantes e irrelevantes, y así poder visualizar los conceptos como una descripción preliminar de las necesidades de información del usuario.
- 4. Feedback:** Puesto que el contenido del documento y las necesidades de información del usuario son representadas imperfectamente por los sistemas actuales de I.R, una recuperación inicial raramente entrega el documento deseado por el usuario, por lo tanto, se realizan varias iteraciones modificando la consulta para alcanzar resultados aceptables. Estas modificaciones se pueden hacer explícitamente por el usuario si decide ingresar una nueva consulta o modificar la original. Por otro lado, el usuario puede comunicarse simplemente con el sistema e indicarle qué documentos desea recuperar, para que así implícitamente el sistema actualice la consulta. Este proceso se denomina Relevance Feedback (RF).

Indicar a los sistemas de I.R cuáles del conjunto de documentos recuperados es relevante o no a sus necesidades, puede servir para clasificarlos como documentos positivos y negativos para un aprendizaje supervisado. El aprendizaje en el contexto del RF tiene la ventaja que una descripción inicial de un concepto para las necesidades de información del usuario puede ser derivada de una consulta.

Dentro del proceso de I.R hay muchas posibilidades de aprendizaje. Una diferencia útil es entre el aprendizaje a corto plazo dentro de una única sesión de búsqueda y el de largo plazo producto de muchas sesiones. Un ejemplo de aprendizaje de corto plazo es

inferir sobre las necesidades de información del usuario a través de la consulta generada, o utilizar el aprendizaje supervisado para refinar los procedimientos que diferencian entre documentos relevantes e irrelevantes.

Un ejemplo de aprendizaje a largo plazo consiste en adquirir y mejorar los sistemas de conocimiento permanente, cambiando la representación de los documentos en respuesta a los repetidos fracasos en la recuperación, o desarrollando un perfil de las necesidades de información y preferencias de un usuario en particular.

En Belkin [15], se destacan las tres alternativas más usadas para aplicar I.R, que son los modelos Booleanos, los vectores de espacio y los modelos de recuperación probabilísticos.

El primero se basa sobre el principio del “*coincidencia exacta*” y los otros dos sobre el concepto de la “*mejor coincidencia*”.

La recuperación Booleana (Bolean Retrieval) se basa sobre el concepto de una coincidencia exacta de una consulta específica con uno o más textos. El término booleano se utiliza porque la consulta se expresa con palabras o frases que combinan el uso de los operadores estándares de la lógica booleana. En este modelo todos los textos que contienen la combinación de frases o palabras especificados en la consulta, son recuperados y no se hacen diferencias entre cualquiera de éstos. Así, el resultado de la operación es una partición de la base de datos entre un conjunto de documentos recuperados y un conjunto de documentos no recuperados.

El modelo de Recuperación Booleana, es el modelo estándar para sistemas de recuperación de información operacional a gran escala. Un problema de este modelo es que no permite hacer un ranking de relevancia del conjunto de documentos recuperados. Los documentos resultantes presentados a los usuarios con algún orden de relevancia dan lugar a sistemas más seguros y utilizables.

Para solucionar los problemas de los modelos de coincidencia exacta, se proponen los modelos de recuperación mejor coincidencia. El más conocido de estos modelos es el Vector de Espacios, este modelo trata textos y consultas como vectores en un espacio multidimensional, las dimensiones son las palabras usadas para representar los textos. Las consultas y textos son comparados a través de un vector de comparación, usando por ejemplo las medidas de similitud de la correlación del coseno. La hipótesis es que cuanto más parecido sea un vector de texto a un vector de consulta, más probable es que el texto sea relevante para la consulta.

En este modelo se hace un refinamiento importante con respecto a qué términos o dimensiones de una consulta, o la representación del texto, pueden ser pesadas (weight), es decir, asignar un valor que determine su importancia dentro del documento. Estos pesos son procesados en base a las distribuciones estadísticas de los términos en la base de datos y en el texto.

Los modelos de recuperación probabilístico se basan sobre el principio de ranking de probabilidad. Esto nos dice que la función de un sistema de I.R es rankear los textos en la base de datos, según la importancia de su probabilidad en la consulta. Este principio toma en cuenta que la necesidad de representación de información y texto es incierta, y la relevancia de las relaciones entre ellas es también incierta.

Este modelo sugiere que hay una variedad de formas de demostración que podrían utilizarse para estimar la probabilidad de relevancia de un texto para una consulta. La forma más típica de demostración es a través de la distribución estadística de términos en la base de datos y en los textos relevantes y no relevantes.

Otro modelo utilizado para I.R es la red de inferencias que se basa sobre redes de inferencia bayesiana. Éstas son grafos dirigidos con dependencia acíclica, en los cuales los nodos representan variables proposicionales o constantes y los arcos representan relaciones de dependencia entre las proposiciones. Si una proposición representada por el nodo p causa o implica la proposición representada por el nodo q , se dibuja un arco directo de p a q . El nodo q contiene una matriz que especifica la $P(p/q)$ para todos los posibles valores de las dos variables.

La representación de texto, categorización de texto, o indexación, ha sido uno de los focos principales de la investigación en I.R ya que se utiliza para clasificar historias de noticias, para buscar información relevante sobre el W.W.W y para guiar las búsquedas de los usuarios a través del hipertexto, según T. Joachims [16], es decir, es el proceso de agrupar documentos dentro de diferentes categorías o clases. Esta técnica cuando se combina con modelos apropiados de recuperación, es asombrosamente eficaz, así como eficiente y de implementación directa. Indexar un objeto para filtrar su información con la utilización de este método, consiste en realizar una exploración léxica para identificar las palabras, haciendo un análisis morfológico para reducir diferentes formas de palabras para obtener descendientes comunes, y contar las ocurrencias de estas descendencias. La simplicidad de este proceso significa que los métodos probabilísticos para la filtración de información son factibles, incluso con volúmenes muy altos de datos entrantes.

T. Joachims [16] en su trabajo utiliza la categorización de texto apoyada por mecanismos de vectores (Supporting Vector Machines, SVMs), además define la categorización de texto como la clasificación de documentos dentro de un número fijo de

categorías predefinidas, donde cada documento puede estar en múltiples, en sólo una o en ninguna de las categorías. Dentro de este trabajo el uso de máquinas de aprendizaje, tiene por objetivo que los clasificadores de texto aprendan desde los conjuntos de documentos que ejecutan las asignaciones de categorías automáticamente, ya que construir un clasificador de texto en forma manual es una tarea difícil y consume demasiado tiempo, este aprendizaje es supervisado.

El primer paso en la categorización de texto es transformar los documentos, que generalmente son string de caracteres, a una representación adecuada para el algoritmo de aprendizaje y las tareas de clasificación. Para ello I.R utiliza la idea de representar las palabras como unidades, es decir, a través de “*Stem*”, que es la raíz de una palabra, por ejemplo la raíz o stem de la palabra computadores es computador. Esto nos lleva a dar valor a estos atributos representados en el texto. Cada stem w_i diferente corresponde a una función, cuyo valor es el número de veces que el stem w_i aparece en el documento. Para evitar vectores innecesariamente largos, las palabras se consideran importantes sólo cuando aparecen por lo menos tres veces y si estas no son stop words.

Un stop word es una palabra que da sentido a un texto, es decir son conjunciones, artículos, preposiciones, etc.

Los mecanismos de apoyo a través de vectores (SVMs) se basan en el principio de minimización del riesgo estructural de la teoría de aprendizaje computacional, su idea es encontrar una hipótesis h para la cual se pueda garantizar el error mínimo. El mínimo error de h es la probabilidad que h sea un error inadvertido en un documento de prueba seleccionado al azar.

En su forma básica, los SVMs tienen un aprendizaje lineal de funciones “*threshold*”. Sin embargo, por un simple “*plug-in*” de una apropiada función kernel, pueden ser utilizados para el aprendizaje de clasificadores polinómicos, redes básicas de función radial (RBF), y redes neuronales “*sigmoid*” de 3 capas.

Una característica notable de los SVMs es que su habilidad para aprender puede ser independiente de la dimensionalidad del espacio. Los SVMs miden la complejidad de la hipótesis basándose en el margen sobre el cual ellos separan los datos, no el número de funciones. Esto significa que se puede generalizar incluso en presencia de muchas funciones, si los datos son separables con un margen amplio usando funciones desde el espacio de la hipótesis.

Si bien este trabajo muestra que las SVMs son una buena opción para utilizarlos como técnicas dentro de la categorización de texto, es bastante vago en su explicación de la forma en que éstos puede ser obtenidos, y cómo se realiza el proceso de categorización de

texto, por lo tanto este trabajo se considera más bien informativo en cuanto a la utilidad y los buenos resultados que puede dar el uso de los SVM_S en la categorización de texto, pero no da una idea clara si su utilidad pueda ser extendida para resolver el problema que se estudia en este proyecto, que es la integración de documentos semánticamente distintos, donde se busca encontrar el grado de similitud entre las frases que serán comparadas.

2.5.1 Algoritmo Rocchio

El algoritmo Rocchio es uno de más populares y el más utilizado en Information Retrieval como método de aprendizaje.

La aplicación de este algoritmo es estudiado en T. Joachims [17], en el cual se analiza el algoritmo Rocchio específicamente en el campo de la categorización de texto, enfocándose principalmente en el peso de las palabras dentro de un documento y las métricas de similitud. También se incluyen mejoras que llevan a una variante probabilística del clasificador Rocchio.

El algoritmo Rocchio inicialmente se desarrolló para la optimización de consultas de Relevance Feedback, el algoritmo se puede adaptar a la categorización de texto y a problemas de enrutamiento.

El término Relevance Feedback, se asocia a tomar los resultados iniciales devueltos por una consulta y utilizar dicha información que puede ser relevante o no en la generación de una nueva consulta.

El mayor componente heurístico del algoritmo Rocchio es el clasificador TFIDF (Term Frequency / Inverse Document Frequency), basado en algoritmos de Relevance Feedback, para modelos de recuperación de vectores de espacio.

Dentro de T. Joachims [17] se presenta una modificación del algoritmo de clasificación TFIDF, llevando al algoritmo a una versión probabilística denominada, PrTFIDF, que optimiza el algoritmo TFIDF y las pruebas realizadas muestran que no sólo da una mejor comprensión teórica del algoritmo TFIDF sino que también se ejecuta mejor en la práctica sin ser más complejo conceptual o computacionalmente. Los puntos 2.5.2, 2.5.3 y 2.5.4 describen como pueden ser obtenidos cada uno de los clasificadores mencionados.

En T. Joachims [17], la categorización de texto se formaliza de la siguiente manera. Ya que la categorización de texto es la clasificación de documentos en un número fijo de categorías, se asume que cada documento d se asigna a una única categoría. Para darle una mayor formalidad se asume que existe un conjunto de clases C y un conjunto de

documentos D . Además se define la siguiente función $T: D \rightarrow C$, que mapea los documentos a una clase.

$T(d)$ son los documentos en el conjunto de entrenamiento. A través del aprendizaje supervisado la información contenida en el conjunto de documentos de entrenamiento se puede utilizar para encontrar un modelo o hipótesis, definiendo la función $H: D \rightarrow C$, que se aproxima a T . $H(d)$ es la función que define las clases a la cual la hipótesis de aprendizaje asigna un documento d y se puede utilizar para clasificar nuevos documentos. El objetivo es encontrar una hipótesis que maximice la exactitud, por ejemplo, la cantidad de veces que H y T concuerdan.

Para poder utilizar estas funciones cada documento que se encuentra como un string de caracteres, se tiene que transformar a una representación adecuada para los algoritmos de aprendizaje y tareas de clasificación. Dentro de I.R se sugiere que las palabras trabajan mejor si se consideran como representación de unidades y que el orden de éstas en un documento no es relevante para las tareas de clasificación, por lo tanto, esta sugerencia lleva a una representación de documentos como bolsa de palabras (bags of words). Esta representación es equivalente a una representación de atributo-valor usado en los mecanismos de aprendizaje. Cada una de las palabras corresponde a una función con el número de ocurrencias de ésta dentro del documento (como valor). Para evitar vectores característicos muy grandes las palabras sólo se consideraran representativas si ocurren en el documento más de 3 veces, este vector se denomina F . La figura 2.18 muestra un ejemplo de un vector para un documento dado.

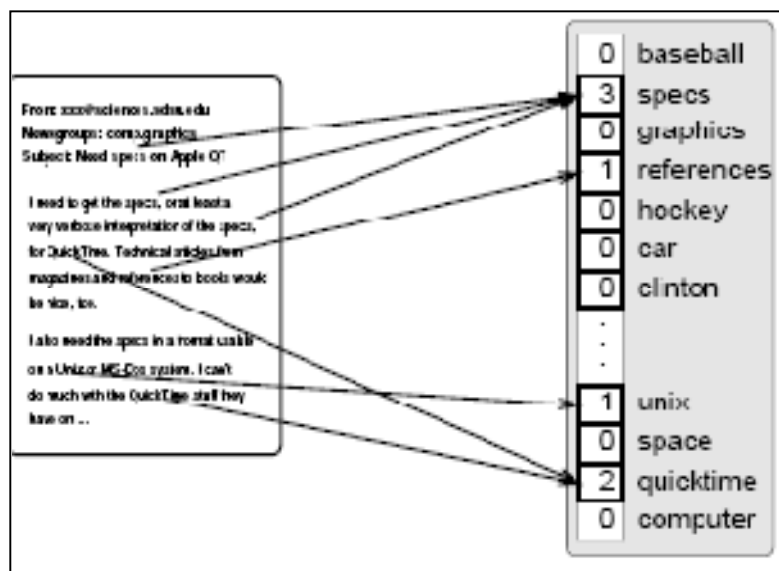


Figura 2.18. Representación de un vector para un documento

2.5.2 Clasificador TFIDF

Este clasificador se basa en el algoritmo de Relevance Feedback propuesto originalmente por Rocchio para el modelo de recuperación de vector de espacio. Debido a sus componentes heurísticos existen tres opciones de diseño que son:

- El método de pesos de palabras.
- Normalización de la longitud del documento.
- Medidas de similitud.

Para obtener los pesos de las palabras se utilizará la técnica de Term Frequency (TF), es decir, la frecuencia de una palabra en un documento, para normalizar la longitud del documento se usará la longitud de vector Euclidiano y para mediar la similitud una función Coseno.

Para obtener este clasificador se hacen las siguientes consideraciones mencionadas en T. Joachims [17]. Cada documento d es representado como un vector $\vec{d} = (d^{(1)}, \dots, d^{(F)})$, por lo tanto, documentos con contenido similar tienen vectores similares (de acuerdo a la medida de similitud fijada). Cada elemento $d^{(i)}$ representa palabras w_i diferentes. $d^{(i)}$ es calculado como una combinación estadística $TF(w_i, d)$ y $IDF(w_i)$, donde Term Frequency $TF(w_i, d)$ es el número de veces que la palabra w_i ocurre en el documento d y el Document Frequency $IDF(w_i)$, es el número de documentos en la cual la palabra w_i ocurre al menos una vez. Por lo tanto, el peso de cada palabra $d^{(i)}$ se puede calcular de la siguiente manera:

$$d^{(i)} = TF(w_i, d) \cdot IDF(w_i) \quad (2.6)$$

El Inverse Document Frequency IDF se puede calcular de la siguiente manera:

$$IDF(w_i) = \log\left(\frac{|D|}{DF(w_i)}\right) \quad (2.7)$$

Con $|D|$ el total de documentos. El IDF será bajo si la palabra se encuentra en muchos documentos pero será alto si ésta se encuentra en sólo uno.

Para lograr el aprendizaje se combinan vectores de documentos en un prototipo de vector \vec{c}_j para cada una de las clases C_j . Primero se normalizan tanto los vectores que contienen los documentos positivos y los vectores que contienen el conjunto de documentos negativos, es decir, los documentos que tienen similitud con la consulta realizada y los que

no. El vector se calcula como una diferencia ponderada de ambos como muestra la ecuación 2.8.

$$\vec{c}_j = \alpha \frac{1}{|C_j|} \sum_{\vec{d} \in C_j} \frac{\vec{d}}{\|\vec{d}\|} - \beta \frac{1}{|D - C_j|} \sum_{\vec{d} \in D - C_j} \frac{\vec{d}}{\|\vec{d}\|} \quad (2.8)$$

α y β son parámetros que se ajustan al conjunto de documentos positivos y negativos y son 8 y 16 respectivamente, estos valores fijos acentúan la ocurrencia tanto en documentos relevantes como no relevantes.

C_j es el conjunto de documentos asignados a la clase j y $\|\vec{d}\|$ denota la longitud Euclidiana de un vector \vec{d} .

El conjunto de vectores resultantes, es un vector para cada una de las clases, lo que representa el modelo de aprendizaje.

Si se desea clasificar un nuevo documento d' . Nuevamente éste se representa como un vector \vec{d}' usando el modelo descrito anteriormente. Para clasificar d' los cosenos de los vectores prototipo de \vec{c}_j con \vec{d}' son calculados. d' se asignará a la clase con la cual su vector de documento tenga el coseno mayor.

$$H_{TFIDF}(d') = \operatorname{argmax}_{C_j \in C} \cos(\vec{c}_j, \vec{d}') \quad (2.9)$$

$\operatorname{argmax} f(x)$ devuelve el argumento x para el cuál $f(x)$ es máximo y $H_{TFIDF}(d')$ es la categoría a la cual el algoritmo asigna documentos d' . El algoritmo se puede resumir de la siguiente manera:

$$H_{TFIDF}(d') = \operatorname{argmax}_{C_j \in C} \frac{\sum_{i=1}^{|F|} c_j^{(i)} \cdot d'^{(i)}}{\sqrt{\sum_{i=1}^{|F|} (c_j^{(i)})^2}} \quad (2.10)$$

2.5.3 Clasificador Naive Bayes

Este clasificador usa un modelo probabilístico de texto según lo descrito en T.Joachims [17]. Para la elaboración de este clasificador se usa un modelo texto *unigram* basado en palabras, es decir, palabras que se supone ocurren independientemente de las otras palabras en el documento. Hay $|C|$ de estos modelos, uno para cada categoría. Todos los documentos asignados para una categoría en particular se suponen que serán generados de acuerdo al modelo asociado con esta categoría.

Para estimar la probabilidad que un documento d' se encuentre en una clase C_j se utiliza una estimación $Pr(C_j|d')$. Las reglas de Bayes dicen, para que d' alcance una exactitud alta en la clasificación se debería asignar a la clase para la cual $Pr(C_j|d')$ es más alta.

$$H_{BAYES}(d') = \operatorname{argmax}_{C_j \in C} Pr(C_j|d') \quad (2.11)$$

$Pr(C_j|d')$ puede ser dividida considerando documentos por separado de acuerdo a su longitud l , donde:

$$Pr(C_j|d') = \sum_{l=1}^{\infty} Pr(C_j|d', l) \cdot Pr(l|d') \quad (2.12)$$

$Pr(l|d')$ es igual a 1 para la longitud l del documento d' y es 0 en otro caso. Aplicando el teorema de Bayes para $Pr(C_j|d', l)$ que es la probabilidad de encontrar un documento d' en la clase C_j dada su longitud l , la ecuación anterior se puede escribir como:

$$Pr(C_j|d') = \frac{Pr(d'|C_j, l') \cdot Pr(C_j|l')}{\sum_{C' \in C} Pr(d'|C', l') \cdot Pr(C'|l')} \quad (2.13)$$

$Pr(C_j|l')$ indica la probabilidad que un documento de longitud l' se encuentre en la clase C_j . Para calcular esta probabilidad se supone, que la categoría de un documento no depende de su longitud, por lo tanto, se puede decir que $Pr(C_j|l') = Pr(C_j)$, entonces se puede calcular un estimador $\hat{Pr}(C_j)$ para $Pr(C_j)$, desde la fracción de documentos que son asignados a la clase C_j .

$$\hat{Pr}(C_j) = \frac{|C_j|}{\sum_{C' \in C} |C'|} = \frac{|C_j|}{|D|} \quad (2.14)$$

$|C_j|$. indica el número de documentos en la clase C_j . y $|D|$ es el número total de documentos.

Dado que el modelo *unigram* como se mencionó anteriormente dice que la ocurrencia de una palabra depende únicamente de la clase de documentos a la que pertenezca, pero su ocurrencia es independiente de las otras palabras presentes en el documento y de la longitud de éste, por lo tanto, $Pr(d'|C_j, l')$ se puede representar de la siguiente manera:

$$Pr(d'|C_j, l') \approx \prod_{i=1}^{|d'|} Pr(w_i|C_j) \quad (2.15)$$

w_i es el rango de la secuencia de palabras en el documento d' , el cuál es un elemento de la función de vectores F . $|d'|$ es el número de palabras en el documento d' . La estimación de $Pr(d'|C_j)$ se reduce a la estimación de cada una de las $Pr(w_i|C_j)$ en forma independiente. Se utiliza un estimador bayesiano para calcular $Pr(w_i|C_j)$:

$$Pr(w_i|C_j) = \frac{1 + TF(w_i, C_j)}{|F| + \sum_{w' \in |F|} TF(w', C_j)} \quad (2.16)$$

$TF(w_i, C_j)$ es el número total de veces que la palabra w ocurre en un documento de la clase C_j .

2.5.4 Clasificador PrTFIDF

El algoritmo PrTFIDF es equivalente a un clasificador TFIDF pero en un entorno probabilístico usando las siguientes consideraciones:

- El mecanismo de ponderación de las palabras utiliza una refinación del peso IDF adaptado especialmente para la categorización de texto.
- La normalización de la longitud del documento se hace usando el número de palabras.
- La medida de similitud es un producto interno.

Este algoritmo hace una estimación de $Pr(C_j|d)$ basándose en la recuperación con indexación probabilística. Donde se utilizan un conjunto de descriptores X para representar el contenido de los documentos. Un descriptor x es asignado a un documento d con una cierta probabilidad $Pr(x|d)$, usando el teorema de probabilidad total y el teorema de Bayes, se tiene que:

$$Pr(C_j|d) \approx \sum_{x \in X} Pr(C_j|x) \cdot Pr(x|d) \quad (2.17)$$

La validez de este supuesto depende de la tarea de clasificación y la elección del conjunto de descriptores X . El descriptor x proporciona bastante información sobre d , la información que no se disponga del documento d se puede obtener tomando en cuenta su categoría de clase C_j .

Una buena forma de escoger X es considerar todos los conjuntos con n palabras del conjunto de funciones F (vector) como potenciales descriptores. El número n de palabras es un parámetro que controla la calidad de la aproximación versus la complejidad de la estimación.

Para obtener la regla de decisión para PrTFIDF, se considera el caso más simple con $n=1$, que da como resultado:

$$\Pr(C_j|d) \approx \sum_{x \in X} \Pr(C_j|w) \cdot \Pr(w|d) \quad (2.18)$$

La probabilidad $\Pr(w|d)$ se puede estimar de la representación del documento d .

$$\Pr(w|d) = \frac{TF(w, d)}{\sum_{w^i \in F} TF(w_i, d)} = \frac{TF(w, d)}{|d|} \quad (2.19)$$

$|d|$ indica el número de palabras en el documento d . $\Pr(C_j|w)$ es la probabilidad que C_j sea la categoría correcta de d , dado que la palabra w del documento d se conoce en forma aleatoria. Una fórmula de Bayes se puede utilizar para reescribir $\Pr(C_j|w)$.

$$\Pr(C_j|w) = \frac{\Pr(w|C_j) \cdot \Pr(C_j)}{\sum_{C' \in C} \Pr(w|C') \cdot \Pr(C')} \quad (2.20)$$

Como se dijo anteriormente, $\Pr(C_j)$ se puede estimar de la parte de documentos que son asignados a la clase C_j como se indica en la fórmula 2.14.

Finalmente $\Pr(w|C_j)$ se puede calcular de la siguiente manera:

$$\Pr(w|C_j) = \frac{1}{|C_j|} \sum_{d \in C_j} \hat{\Pr}(w|d) \quad (2.21)$$

La regla de decisión para PrTFIDF es:

$$H_{PrTFIDF}(d') = \underset{w \in F}{\operatorname{argmax}}_{C_j \in C} \sum_{C' \in C} \frac{\Pr(w|C_j) \cdot \Pr(C_j)}{\Pr(w|C') \cdot \Pr(C')} \cdot \Pr(w|d') \quad (2.22)$$

2.5.5 Clasificación Automática de Documentos

En C. Goller [19], se define la clasificación de documentos como la asignación basada en documentos de una o más categorías de documentos predefinidas. Ésta se puede utilizar para el filtrado de documentos y mecanismos de enrutamiento para el procesamiento de temas específicos tales como extracción de información y traducción automática.

En el proceso de clasificación automática de documentos se distinguen dos fases. Primero la fase de aprendizaje, donde los usuarios definen las categorías o temas en los que ellos están interesados, por lo tanto, existe una necesidad de información, dando un espacio muestral de documentos para cada una de estas categorías, la figura 2.19 muestra esta fase.

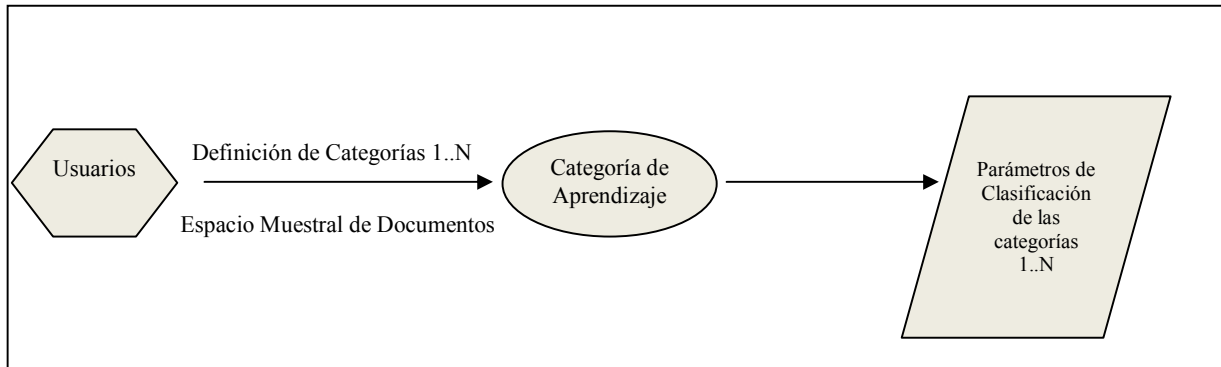


Figura 2.19. Representación de la Fase de Aprendizaje.

La segunda fase es clasificación en la cual se pueden entregar nuevos documentos al Clasificador de Categorías, éste devuelve una Asociación de Categorías, que es una clasificación para cada una de las categorías. Básicamente dice al usuario si el documento se ocupa del entrenamiento de las categorías y específicamente de cuál de ellas se ocupa, por lo tanto, se asume que el conjunto de documentos tiene que ser clasificado. La figura 2.20 muestra la ejecución de esta fase y la figura 2.21 muestra el proceso completo de clasificación de documentos.

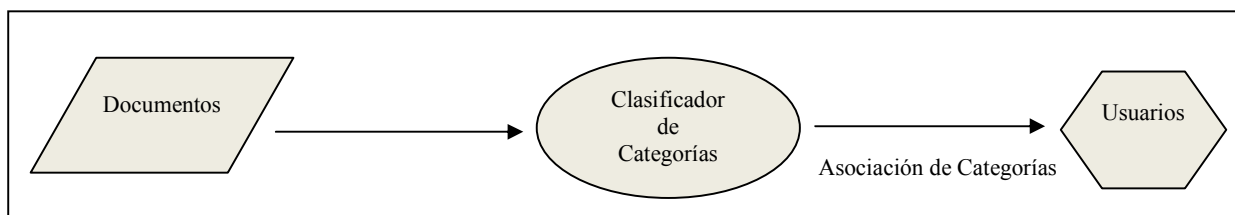


Figura 2.20. Representación de la Fase de Clasificación.

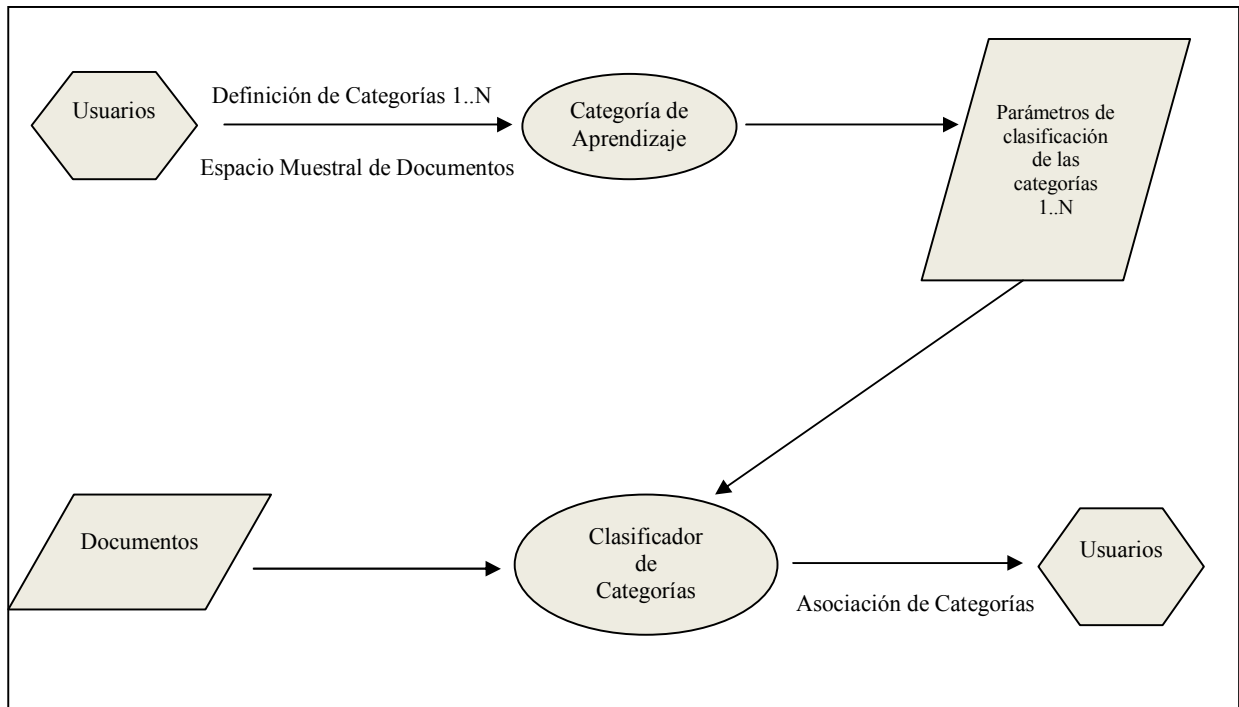


Figura 2.21. Representación de la Clasificación de Documentos.

Dentro de la figura 2.21 el componente Categoría de Aprendizaje analiza los documentos del espacio muestral e intenta identificar el contenido específico de cada una de las categorías. Este análisis es un proceso de razonamiento inductivo que incluye la generalización y abstracción. La salida es un modelo para cada una de las categorías la cual es representada por un conjunto de parámetros de clasificación. La mayoría de los clasificadores hacen supuestos previos sobre el modelo y su complejidad. Si los supuestos de complejidad del modelo son demasiado altos, significa que puede ocurrir un *overfitting*. Esto significa que el modelo está sobre especializado con respecto a los documentos del espacio muestral.

En C. Goller [19], se hace una relación entre la clasificación automática de documentos y la recuperación de documentos. En la recuperación de documentos, los usuarios especifican sus necesidades de información generando una consulta, la que es analizada y posteriormente aplicada a un conjunto de documentos relativamente fijos y reprocesados. Los documentos con mejor *matching* de acuerdo a la consulta son presentados al usuario. La figura 2.22 muestra el proceso de recuperación de documentos.

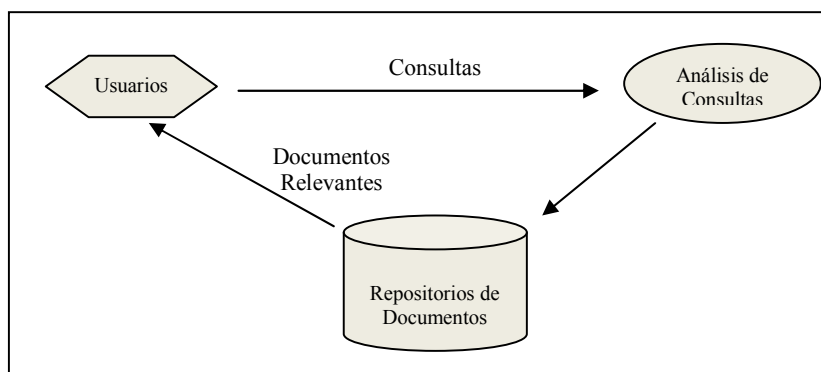


Figura 2.22. Representación de la Recuperación de Documentos.

Hay dos diferencias principales entre los procesos antes mencionados. La Primera es más bien técnica, en el proceso de recuperación los documentos comparados dada la consulta del usuario generalmente pertenecen a una recopilación de documentos grande y relativamente fija, los cuales son preprocesados independientemente de las consultas individuales. Normalmente las consultas son pequeñas y pueden ser aplicadas a toda la recopilación de documentos en forma paralela y no secuencialmente a cada documento. El análisis de consultas tiene que ser muy eficiente, dada la cantidad de preguntas que se responden y las que continuamente llegan.

En cambio, en la clasificación de documentos, las necesidades de información del usuario son relativamente estables. Sin embargo, no tienen agrupaciones fijas de documentos que puedan ser preprocesados. Normalmente, los documentos tienen que ser procesados secuencialmente por el clasificador.

La segunda diferencia se refiere a la manera en como las necesidades de información son especificadas. En la recuperación de documentos se especifican a través de pequeñas consultas (palabras claves). Los usuarios tienen que saber exactamente en que están interesados. La extensión de la consulta usando modelos del dominio es muy importante, dadas las diferencias de lenguaje usadas entre los autores de documentos y los usuarios que especifican la consulta que puede llevar constantemente a una baja memoria.

En la clasificación de documentos los usuarios especifican sus necesidades de información a través de la selección de documentos usando ejemplos y contraejemplos para un tema. Si hay suficientes documentos en la muestra y si éstos son representativos para la clase de documentos encontrados, la clasificación será buena. Los usuarios no tienen que saber lo que buscan exactamente y el uso de modelos del dominio no son necesarios.

Estas diferencias desaparecen, si se considera la recuperación de documentos con un feedback con el usuario, ya que se permite que éstos puedan refinar sus consultas

originales, seleccionando los documentos relevantes y no relevantes desde el conjunto de documentos entregados por los sistemas de recuperación como respuesta a la consulta original.

Para aplicar los métodos de aprendizaje y clasificación se requiere la construcción de un vector característico, en C. Goller [19] se compararon dos técnicas para construir el vector, Letter 5-Gram y un análisis morfológico.

Letter 5- Gram es una técnica que divide cada palabra en diferentes conjuntos de cinco letras, incluye caracteres especiales como “!, %, &, \$” y el espacio en blanco, además de no distinguir entre mayúsculas y minúsculas. Su implementación es muy fácil e independiente del lenguaje, una de sus características es que ejecuta automáticamente ciertas clases de stemming que son útiles para los errores ortográficos.

El análisis morfológico combina la procedencia variable y derivacional de las palabras, junto con un análisis compuesto de la palabra. Luego de la tokenización, es decir, identificar cuáles son palabras dentro del texto y cuáles frases, reconoce los límites de las sentencias, los tokens (símbolos) son analizados como concatenaciones a través de cuatro categorías morfológicas: prefijos, descendencia de la palabra (word stem), sufijos y descendencia solitaria (solitary stem) esto significa que la palabra no es ni prefijo ni sufijo.

Para comparar las técnicas mencionadas se utilizó una conjunto de prueba obtenido de un subconjunto de artículos del diario alemán “Süddeutsche Zeitung (1998)” (SZ), específicamente de la sección Múnich donde se tomó un grupo de 20 categorías con el número más alto de artículos asociados.

El número de palabras únicas en el conjunto de documentos analizados determinan la dimensión del vector que se obtendrá por cualquiera de las dos técnicas antes mencionadas, y la posición de cada una de las palabras en una lista ordenada alfabéticamente determina su posición dentro del vector, por lo tanto el vector de palabras es simplemente un vector de pesos.

Este trabajo hace una buena definición de lo que es la clasificación de texto junto con mostrar las fases asociadas a este proceso. Además deja clara la diferencia con respecto a la recuperación de documentos. Todo esto amplía los conocimientos adquiridos. Pero en cuanto a las técnicas de análisis morfológico y letter 5-gram, se puede decir que la primera es más aplicable al área de estudio, que es poder hacer una comparación de competencias académicas obtenidas durante un período de estudio, donde el volumen de información es grande.

Capítulo 3

Iniciativas de Integración de los Sistemas de Certificación a Nivel Europeo

3.1 Introducción

Uno de los obstáculos principales para la gente que desea trabajar o aprender en otro país de la U.E, “*Unión Europea*”, o moverse por diversas partes del mercado laboral europeo, es que sus cualificaciones y competencias no son aceptadas. Esto se complica aún más por la proliferación de las cualificaciones en todo el mundo, la diversidad de los sistemas nacionales de cualificación y de las estructuras de educación y entrenamiento, y el constante cambio en estos sistemas.

La validación del «aprendizaje no formal» también es importante, ya que éste puede adquirirse, por ejemplo, en el lugar de trabajo o mediante una actividad voluntaria, convirtiéndose en un elemento fundamental del aprendizaje permanente, que permite a los ciudadanos la adquisición de conocimientos y competencias dondequiera que sea posible.

Otra forma de posibilitar la movilización de estudiantes y profesionales es a través del reconocimiento de los títulos profesionales. En este ámbito hay que hacer una distinción clara sobre los propósitos para los cuales se reconocerá el diploma, ya que éste puede adquirir un reconocimiento para efectos académicos, es decir, para que el poseedor continúe sus estudios o puede ser para propósitos profesionales, es decir, que el título sea reconocido para que el individuo se pueda desarrollar en una cierta profesión.

Para afrontar estos obstáculos, la U.E ha introducido varios instrumentos, que tienen como objetivo facilitar la transparencia de cualificaciones y competencias para propósitos académicos o profesionales. Con esto se quiere hacer de Europa, un área de aprendizaje permanente, para permitir a los ciudadanos satisfacer los desafíos de una sociedad basada en el conocimiento, promoviendo el desarrollo de éste y sus competencias en todas las etapas de sus vidas, posibilitándolos para que utilicen sus cualificaciones y competencias como una especie de moneda común, que pueda obtenerse en un sitio y gastarse en otro.

Los programas europeos Leonardo da Vinci y Sócrates/Erasmus han representado un papel clave en el apoyo del aprendizaje permanente y el desarrollo de la formación profesional y la formación transnacional. Estos programas se describirán a continuación en este capítulo.

Otro instrumento destinado a mejorar la transparencia de las cualificaciones para efectos académicos es el “*Sistema Europeo de Transferencia y Acumulación de Créditos*” (“European Credits Transfer and Accumulation System” ECTS), creado por la comisión europea hace más de diez años como base común para el reconocimiento de los periodos de estudios de los estudiantes en el extranjero. Este sistema sigue extendiéndose, incluso fuera de Europa.

Para facilitar la transferencia y transparencia de las cualificaciones y competencias a nivel de reconocimiento profesional, se creó una red de puntos nacionales de referencia, distribuidos en toda la U.E, que es el primer punto de contacto para cuestiones relacionadas con las cualificaciones profesionales. Todos los Puntos Nacionales de Referencia forman parte de una red coordinada que cumple las siguientes funciones.

- Es el primer punto de contacto a escala nacional para asuntos relativos a los Títulos de formación profesional, Certificados de Profesionalidad, Cualificaciones Profesionales, así como para los Suplementos Europass al Título/Certificado y al Título de Técnico Superior de Formación Profesional;
- Ofrece la información pertinente y sirve como punto de contacto con los organismos nacionales que disponen de la información;
- Actúa como asociado nacional dentro de la Red Europea de Puntos Nacionales de Referencia.

3.2 European Qualification Framework (EQF)

La comisión europea para la acreditación en educación superior se estableció para alcanzar el mutuo reconocimiento en las decisiones de acreditación entre los países de la comunidad. El desarrollo de un EQF contribuirá a este objetivo.

El EQF [9] debe contribuir a la comprensión y a la transparencia de la educación superior en Europa y al reconocimiento de cualificaciones. Los objetivos principales del EQF son [8]:

- Facilitar la movilidad de estudiantes y trabajadores.

- Proporcionar información clara a todos los involucrados en el proceso de movilidad laboral y estudiantil con respecto a la naturaleza de las cualificaciones y competencias dentro del framework Europeo.

El framework debe contener al menos los siguientes aspectos:

- Una completa estructura de la educación superior, con ciclos y niveles.
- La reestructuración de la educación superior en Europa comprende ahora tres ciclos principales: el primer ciclo corresponde al nivel “*Bachelor*”, el segundo ciclo al nivel “*Master*” y el tercer ciclo al nivel de “*Doctor*” en un área determinada. Dentro de estos tres ciclos un número de niveles consecutivos y paralelos se puede definir dependiendo del framework nacional. Se considera sin embargo que el nivel más alto dentro de cada ciclo debe dar el acceso al siguiente ciclo.
- Describir las cualificaciones en base a los resultados esperados del aprendizaje. Se han hecho varios modelos en Europa para describir cualificaciones particulares basadas en el aprendizaje resultante, descritos como genéricos, orientado a los temas de estudio y orientado al perfil.
- Generar un sistema de créditos basados en la carga de trabajo académico y las horas de trabajo teóricas. El sistema de transferencia europeo de crédito (ECTS) es un marco extensamente usado para describir las unidades de carga de trabajo.
- El framework de cualificaciones debe proporcionar información sobre los requisitos de acceso para los principales ciclos de cualificación en los sistemas de educación superior (y desde un ciclo a otro) en el país específico. Los requisitos de entrada variarán considerablemente dependiendo de los niveles consecutivos y paralelos dentro de un contexto nacional.
- Define los posibles movimientos desde un programa o ciclo a otro. Por ejemplo, la finalización del primer ciclo permite el acceso al segundo ciclo, etc. Puede definir en una perspectiva más amplia las posibilidades de obtener un trabajo en un área particular del mercado laboral.
- Asegurar la calidad del Framework es fundamental para el establecimiento, la implementación y el mantenimiento exitoso del EQF. El desarrollo de sistemas que aseguren la calidad ocurre en paralelo con el desarrollo del EQF.
- Se acordó en Berlín que el Diploma Suplementario debe estar disponible para todos los graduados en forma gratuita antes del 2005. El diploma suplementario debe informar sobre el ciclo apropiado dentro del EQF.

La creación de un EQF se realizó con el objetivo de dar un primer paso en la estandarización de la educación a nivel Europeo, el framework busca contribuir a transparentar la educación y con ello apoyar la movilización de estudiantes y trabajadores que desean ampliar sus estudios en países diferentes al de origen, ya que su nivel profesional puede ser evaluado a través de este framework común.

A un nivel técnico no universitario poder medir cualificaciones y competencias dentro de este framework es una tarea más difícil, porque éste da sólo una idea general de las competencias y cualificaciones que debe tener una persona en un determinado nivel educacional, por lo tanto esta iniciativa da un mayor apoyo a la transparencia educativa a nivel universitario. La figura 3.1 muestra la propuesta del EQF.

CYCLE	FIRST CYCLE	SECOND CYCLE	THIRD CYCLE
LEVEL	BACHELOR	MASTER	DOCTORAL
Main Awards ²	Bachelor	Master	Doctoral
(Other National Awards) ³			
PROFILE	The knowledge, skill and competence acquired are <i>relevant</i> to personal development, participation in society and community, employment and access to additional education and training. Distinctions <i>might</i> be made between professional/applied and academic/research oriented profiles.	The knowledge, skill and competence acquired are <i>relevant</i> to personal development, participation in society and community, employment, and access to additional education and training. Distinctions <i>might</i> be made between professional/applied and academic/research oriented profiles.	The knowledge, skill and competence acquired are <i>relevant</i> to personal development, participation in society and community, employment, and access to additional education and training. Distinctions <i>might</i> be made between professional/applied and academic/research oriented profiles.
LEARNING OUTCOMES	Dublin Descriptors for generic competences	Dublin Descriptors for generic competences	Under development
ACCESS	Requirements and awards necessary to participate in the specific cycle.	Requirements and awards necessary to participate in the specific cycle.	Requirements and awards necessary to participate in the specific cycle.
PROGRESSION	Progression to second cycle following completion of the first cycle. ⁵	Progression to third cycle following completion of the second cycle.	
WORKLOAD CREDITS based on ECTS	180 to 240	60 to 120	To be determined.

Figura 3.1. Propuesta del EQF.

3.3 Programa Sócrates Erasmus

La enseñanza superior desempeña un papel decisivo en el desarrollo de los ciudadanos y de las sociedades modernas, en la medida que potencia el desarrollo social, cultural, económico, la ciudadanía activa y los valores éticos.

A escala europea, la educación en general, y la enseñanza superior en particular, no son objeto de una política europea común; el contenido y la organización de los estudios sigue siendo competencia de los Estados miembros de la U.E. Sin embargo, conforme al artículo 149 del Tratado de la Comunidad Europea [3], la Comunidad contribuirá al desarrollo de una educación de calidad fomentando la cooperación entre los Estados miembros, a través de una amplia gama de acciones, tales como la promoción de la movilidad de los ciudadanos, la elaboración de planes de estudios comunes, la creación de redes, el intercambio de información o la enseñanza de los idiomas de la U.E.

El Tratado ofrece igualmente un compromiso en favor del aprendizaje permanente para todos los ciudadanos de la Comunidad. En este sentido, la Comunidad debe desempeñar un papel complementario, es decir, dotar a la educación de una dimensión europea, favorecer el desarrollo de una educación de calidad y fomentar el aprendizaje permanente. Las últimas cumbres europeas (a partir de la celebrada en Lisboa en el 2000) han resaltado la contribución de la educación a la creación de la sociedad europea del conocimiento.

El principal instrumento para hacer realidad esta aspiración es el programa Sócrates, que incluye una acción centrada específicamente en la enseñanza superior: Sócrates/Erasmus[10].

Erasmus es la acción destinada a la enseñanza superior del programa Sócrates II. Tiene por objetivo mejorar la calidad y fortalecer la dimensión europea de la enseñanza superior fomentando la cooperación transnacional entre universidades, estimulando la movilidad en Europa y mejorando la transparencia y el pleno reconocimiento académico de los estudios y cualificaciones en toda la Unión.

Su nombre le fue dado en honor al filósofo, teólogo y humanista Erasmo de Rotterdam (1465-1536). Adversario infatigable del pensamiento dogmático en todos los campos del quehacer humano, Erasmo vivió y trabajó en diversas partes de Europa, en búsqueda del conocimiento, la experiencia y las percepciones que sólo ese contacto con otros países podían proporcionarle.

Este programa comprende una gran variedad de actividades: intercambios de estudiantes y profesores, desarrollo conjunto de programas de estudio (Curriculum

Development), programas intensivos internacionales, redes temáticas entre departamentos y facultades de toda Europa, cursos de idiomas (EILC) y el sistema europeo de transferencia de créditos académicos (ECTS).

Los programas de intercambio de estudiantes ofrecen la posibilidad de estudiar en el exterior, es decir, en otro país europeo, por un periodo entre 3 y 12 meses. Los estudiantes reciben un permiso de estudios, y no deben pagar a la universidad de intercambio los costos asociados a su periodo académico. Además estos estudios son reconocidos en la universidad de origen.

Cuando un estudiante decide participar en el programa Erasmus debe firmar un acuerdo escrito entre él, la universidad de origen y la universidad patrocinante, que precisa el programa de estudios que seguirá el alumno en el extranjero. Cualquier cambio a este acuerdo debe ser convenido por las tres partes.

El programa Sócrates/Erasmus no se refiere únicamente a la movilidad de los estudiantes. Por el contrario, incluye un gran número de actividades, especialmente dirigidas a los profesores y demás personal docente de las universidades.

Todas las actividades de Sócrates/Erasmus relacionadas con la enseñanza superior tienen como objetivo generar una dimensión europea en la totalidad de los programas académicos de una universidad. Se ofrece un apoyo financiero a las universidades como incentivo para que introduzcan una perspectiva europea en sus cursos.

Para los profesores y demás personal docente de la universidad, la participación activa en la amplia gama de actividades propuestas, es esencial para acercar su universidad a Europa y Europa a su universidad. Esta experiencia supone un impacto evidente no sólo en el profesor directamente implicado, sino también en los estudiantes de la universidad de origen y de acogida. Por lo tanto, este programa busca mejorar la calidad y fortalecer la enseñanza superior fomentando la cooperación transnacional entre universidades.

Esta iniciativa se realiza con el objetivo de aumentar la movilización de estudiantes, reconociendo sus estudios, pero es aplicada únicamente a nivel universitario. Dado que medir o comparar certificados de estudios y competencias a este nivel es mucho más fácil, ya que las carreras universitarias con respecto a sus planes de estudios no cambian mucho entre una universidad y otra y entre un país y otro, por consiguiente, medir el grado de conocimiento de una persona es más simple. No así para el nivel técnico no universitario, ya que medir el grado de conocimiento o “el saber hacer”, que es lo buscado por este proyecto, requiere de un análisis mucho más detallado que no se puede lograr a nivel de plan de estudios únicamente.

3.4 European Credit Transfer and Accumulation System (ECTS)

Para lograr una comunicación entre las diferentes universidades a nivel de estudiantes que completan estudios en el extranjero se estableció un sistema de créditos a nivel europeo. Un sistema de créditos es una forma sistemática de describir un programa de educación asignando créditos a sus componentes. La definición de los créditos en los sistemas de educación superior puede basarse en distintos parámetros, como la carga de trabajo del estudiante, los cursos y objetivos de formación, los resultados del aprendizaje y las horas de contacto.

El único sistema de crédito que se ha ensayado y utilizado con éxito en Europa, a nivel universitario es, ECTS que es el “*European Credit Transfer and Accumulation System*” [5] es decir, el “*Sistema Europeo de Transferencia y acumulación de Créditos*”, este sistema está orientado al estudiante y se basa en la carga de trabajo, necesaria para alcanzar los objetivos de un programa académico. Estos objetivos se especifican en términos de los resultados del aprendizaje y de las competencias que se van a adquirir.

El ECTS fue adoptado en 1989, en el marco del programa Erasmus. Este sistema facilita el reconocimiento de periodos de estudios en el extranjero, incrementando así la calidad y el volumen de la movilidad de los estudiantes en Europa, porque hace que los programas de estudio resulten fácilmente comprensibles y comparables para todos los estudiantes, tanto locales como extranjeros; facilita la movilidad y el reconocimiento académico; ayuda a las universidades a organizar y revisar sus programas de estudios; además puede ser utilizado por diversos programas y modalidades de enseñanza; haciendo que la educación superior europea sea más atractiva para los estudiantes de otros continentes.

El sistema ECTS tiene las siguientes características esenciales:

- Se basa en la convención de que 60 créditos miden la carga de trabajo de un estudiante a tiempo completo durante un curso académico, que en Europa equivale en la mayoría de los casos a 1500 o 1800 horas por año, y en tales casos un crédito representa de 25 a 30 horas de trabajo.
- Los créditos sólo se pueden obtener una vez que se ha completado el trabajo requerido y se ha realizado la evaluación adecuada de los resultados del aprendizaje. El crédito es también una forma de cuantificar los resultados del aprendizaje. Que son conjuntos de competencias que expresan lo que el estudiante sabrá,

comprenderá o será capaz de hacer tras completar un proceso de estudios, corto o largo.

- La carga de trabajo del estudiante en el ECTS consiste en el tiempo invertido en asistencia a clases, seminarios, estudio personal, preparación y realización de exámenes, etc.
- Se asignan créditos a todos los componentes educativos de un programa de estudios (como módulos, cursos, periodos de prácticas, trabajos de tesis). Los créditos reflejan el volumen de trabajo que cada componente requiere en relación con el volumen total de trabajo necesario para completar un curso.
- Los resultados del estudiante se documentan mediante la atribución de una nota local/nacional. Es una buena práctica añadir un grado ECTS, especialmente en caso de transferencia de créditos. En la escala de grados del ECTS, los resultados de los estudiantes son clasificados sobre una base estadística. Por tanto, los datos estadísticos sobre los resultados del estudiante son un requisito previo para aplicar el sistema de calificación del ECTS. Las notas se asignan entre los estudiantes del siguiente modo: A el 10 % mejor, B el 25 % siguiente, C el 30 % siguiente, D el 25 % siguiente, E el 10 % restante. Se hace una distinción entre grados “FX” y “F”, que se asignan a los estudiantes que no aprueban. “FX” significa: “Suspensión. Se requiere un poco más de trabajo para aprobar”, y “F” significa: “Suspensión. Se requiere un gran trabajo para aprobar”. La inclusión de las tasas de suspensión en el expediente académico es opcional.

Los documentos esenciales que debe contener el ECTS son:

- El catálogo informativo, la guía de cursos de la institución publicado en dos idiomas (o sólo en inglés en el caso de los programas impartidos en inglés) en la web y/o en versión impresa, en un único documento o en varios. El catálogo informativo o guía de cursos debe incluir todos los puntos de la lista de títulos anexa al presente documento, incluida la información dirigida a los estudiantes extranjeros que acoge la institución.
- El contrato de estudios contiene la lista de asignaturas o materias que se habrán de realizar, con los créditos ECTS atribuidos para cada curso, y que habrán de acordar el estudiante y el organismo académico responsable de la institución. En caso de transferencia de créditos, el contrato de estudios ha de ser acordado por el estudiante y las dos instituciones antes de la partida de éste, y ha de ser actualizado inmediatamente cuando hayan ocurrido cambios.
- El expediente académico documenta los resultados del estudiante indicando la lista de asignaturas o materias realizadas y los créditos obtenidos, así como las notas

locales y, en su caso, los grados ECTS otorgados. En caso de transferencia de créditos, la institución de origen expedirá el certificado académico para los estudiantes salientes antes de su partida, y la institución de destino lo hará para los estudiantes que acoge al final de su período de estudios.

El programas Erasmus como el sistema ECTS sólo es adecuado a nivel universitario. A nivel técnico no universitario estas formas de reconocimiento de competencias y habilidades adquiridas por un estudiante que completa estudios en el extranjero, no es el adecuado ya que no asegura que los estudiantes a nivel técnico obtengan el perfil profesional requerido en el país de origen, debido a que este programa no hila fino con respecto al reconocimiento de competencias y contenidos que son fundamentales para este sector profesional, ya que a nivel técnico se deben medir los conocimientos de manera práctica, es decir, en el “saber hacer”.

3.5 Europass

El Europass es un conjunto de documentos que ayuda a los ciudadanos a comunicar de manera clara y sencilla las aptitudes, las titulaciones y certificaciones adquiridas a lo largo de la vida, tanto entre países como entre sectores.

Su principal objetivo [2] es “Facilitar la movilidad de los estudiantes y los trabajadores a través de los Estados Miembros de la U.E, los países del Área Económica de Libre Comercio/Espacio Económico Europeo y los países candidatos, siempre que se desee buscar trabajo o solicitar la admisión en algún programa educativo o formativo”.

El Europass engloba un conjunto de cinco documentos, siendo el Currículum Vitae su principal elemento y pudiendo complementarlo con los restantes documentos Europass, dependiendo del historial de cada individuo.

3.5.1 Currículum Vitae

El Currículum Vitae Europass (CV) [23], pone a disposición de los ciudadanos un modelo común para presentar de modo sistemático, cronológico y flexible sus competencias, titulaciones y certificaciones. El modelo incluye indicaciones específicas sobre los distintos campos, así como una serie de directrices y ejemplos.

Este documento, está disponible en varios idiomas europeos, es un documento personal que contiene información exclusivamente individual. Incluye distintas secciones para la presentación de información sobre datos personales, competencias lingüísticas, experiencia laboral, nivel de estudios y formación. Además de otras competencias adicionales del interesado, haciendo hincapié en las competencias técnicas, organizativas,

artísticas y sociales. La información adicional puede añadirse al CV Europass en forma de uno o varios anexos.

Este documento constituye el principal elemento del Europass: el expediente Europass de un ciudadano incluirá el CV Europass completado por el propio ciudadano y uno o varios documentos Europass más, dependiendo de los conocimientos y el historial laboral específico de cada persona.

3.5.2 Pasaporte de Lenguas Europass

El Pasaporte de Lenguas Europass [24], permite presentar los conocimientos lingüísticos, que se hayan adquirido o no en el marco de la enseñanza formal, este es un aspecto vital a la hora de estudiar y trabajar en Europa, aunque no es un "documento acreditativo oficial", ya que sólo organismos oficiales, pueden acreditar los conocimientos de idiomas mediante una serie de pruebas o exámenes también oficiales.

El pasaporte también está diseñado para ayudar a los alumnos a reflexionar sobre los objetivos de su aprendizaje, a planificarlo y a aprender de manera autónoma. Complementa al CV Europass y puede adjuntarse a éste.

En cuanto a la función informativa, documenta las competencias lingüísticas de su titular de manera exhaustiva, informativa, transparente y fidedigna. Ayuda al titular a dejar constancia del nivel de competencia alcanzado en una o varias lenguas extranjeras y, al mismo tiempo, informa a terceros de forma detallada y comparable internacionalmente.

El Pasaporte de Lenguas Europass, está disponible en varios idiomas Europeos, es un documento personal de libre acceso. Contiene una certificación de idiomas extranjeros que su titular pone al día con regularidad. Donde éste describe sus conocimientos lingüísticos, de acuerdo con criterios comunes aceptados en Europa. Además contiene una biografía lingüística detallada que describe la experiencia del titular en cada lengua y un expediente que permite archivar ejemplos de trabajos personales para ilustrar los conocimientos lingüísticos de éste.

3.5.3 Documento de Movilidad Europass

El Documento de movilidad Europass [25], es un documento personal que registra un periodo de aprendizaje o formación a nivel universitario, realizado por el titular en un país extranjero, permitiéndole comunicar mejor sus experiencias y, en concreto, las competencias adquiridas. Este documento requiere un acuerdo escrito de colaboración entre la organización de envío y la organización que recibe al titular, estableciendo los contenidos, los objetivos, la duración, la metodología durante la permanencia formativa y designando un tutor en la organización de acogida, encargado de ayudar, informar, guiar y supervisar al interesado.

Un individuo no puede solicitar el documento personalmente. Todas las solicitudes deben hacerse a través del centro que organice el periodo de aprendizaje o formación en un país europeo extranjero.

En el caso de los proyectos de movilidad transnacional dentro del marco de los programas comunitarios en el ámbito de la educación y la formación (Leonardo da Vinci, Comenius, Erasmus), la solicitud del documento se debe hacer a través del programa correspondiente. Por el momento, sólo se puede solicitar este documento en el caso de que se trate de un proyecto Leonardo da Vinci.

El documento de Movilidad Europass contiene datos personales, pero de éstos el único campo obligatorio es el nombre de la persona a la que se extiende el documento. Las organizaciones que extienden este certificado sólo pueden rellenar los demás campos relativos a los datos personales si el interesado está de acuerdo.

El campo cualificación tampoco es obligatorio, reconociéndose así el hecho que no todas las iniciativas de educación y formación conducen a una cualificación formal.

3.5.4 Certificate Supplement Europass

El Certificate Supplement [26], es un documento informativo e institucional destinado a fomentar la transparencia y la transferencia de los títulos y certificados profesionales. Sus objetivos son:

- Ampliar la información sobre el Título de Técnico de Formación Profesional o del Certificado de Profesionalidad, en términos de las competencias que desarrolla el titular, la gama de empleos accesible, los organismos de expedición y acreditación, el nivel del título o certificado, las distintas maneras de obtener el título o el certificado, los requisitos de entrada y las posibilidades de acceso al siguiente nivel de formación.
- Facilitar la comprensión a empleadores o instituciones de otro país del significado del título o certificado.

El C.S no sustituye ni da derecho al reconocimiento formal del certificado original por las autoridades académicas de otros países. Por otra parte, facilita una apreciación justa del certificado original, de modo que puede ser útil para obtener el reconocimiento de las autoridades competentes.

Este certificado se utiliza como base para poder hacer una comparación entre los certificados entregados por las diferentes instituciones encargadas de dar un educación

técnica no universitaria, ya que establece los campos mínimos que debe contener un certificado para que pueda realizarse un mapping entre éstos, es decir poder integrarlos con el objetivo de poder medir las competencias adquiridas por un estudiante, que ha realizado estudios en el extranjero.

3.5.5 Suplemento Europass al Título Superior

El Suplemento Europass al título Universitario [27], es un documento informativo y personalizado adjunto al Título Universitario o al Título de Técnico Superior de Formación Profesional, que describe la naturaleza, el nivel, el contexto, el contenido y la normativa de la formación realizada por el titular. Además da amplia información sobre los resultados obtenidos por cada poseedor del certificado y sobre el sistema nacional de enseñanza superior.

El Suplemento no sustituye ni da derecho al reconocimiento o validación formal del título, que se realizará, en su caso, por los procedimientos legislativos establecidos por cada Estado miembro. Sin embargo, tiene como objetivo ampliar la información y facilitar una apreciación justa del título original, de modo que pueda ser útil para obtener el reconocimiento de las autoridades competentes, del personal administrativo de las instituciones de formación superior y de futuros empleadores.

3.6 Programa Leonardo da Vinci

Este programa contribuye a la realización de un espacio educativo europeo que fomenta la educación y el aprendizaje permanente y garantiza el mantenimiento de la cooperación a nivel comunitario entre agentes de la formación profesional.

Los objetivos principales del programa son los siguientes [11]:

- Mejorar las aptitudes y competencias individuales, especialmente de los jóvenes, este objetivo podrá conseguirse mediante la formación profesional y el aprendizaje en alternancia con el trabajo, con la perspectiva de facilitar la inserción profesional.
- Mejorar la calidad de la formación profesional continua y de la adquisición de aptitudes y competencias a lo largo de la vida, así como el acceso a las mismas, con la intención de incrementar y desarrollar la capacidad de adaptación.
- Promover y reforzar la contribución de la formación profesional al proceso de innovación a fin de mejorar la competitividad y el espíritu empresarial.

La aplicación de estos objetivos a escala europea apoya y complementa las políticas de los Estados miembros. La Comisión a cargo del programa velará por la coherencia de las acciones de éste con las demás acciones y políticas de la Comunidad.

Se presta especial atención a las personas desfavorecidas en el mercado del trabajo, incluidas aquellas que tienen alguna discapacidad, así como a las prácticas que faciliten su acceso a la formación, al fomento de la igualdad y a la lucha contra la discriminación.

Para alcanzar los objetivos del programa, se han contemplado las siguientes medidas:

- Apoyo a la movilidad transnacional de las personas que sigan una formación profesional, así como de los responsables de la formación en Europa.
- Apoyo a proyectos piloto basados en asociaciones transnacionales, concebidos para potenciar la innovación y la calidad de la formación profesional.
- Promoción, en el marco de la formación profesional, de las competencias lingüísticas, incluso en los idiomas de menor difusión y enseñanza, así como de la comprensión de las distintas culturas.
- Apoyo al desarrollo de redes de cooperación transnacionales que permitan el intercambio de experiencias y de prácticas idóneas.
- Elaboración y actualización de la documentación comunitaria de referencia y de datos comparables.
- Desarrollo de acciones conjuntas con otros programas comunitarios.

3.6.1 Proyecto B.E.A.T.R.I.C

El proyecto B.E.A.T.R.I.C [1] “*Building european Passport for Trnasporent and International Certification Proyect*”, pertenece al programa Leonardo da Vinci y busca construir un pasaporte europeo internacional para transparentar los procesos de certificación a nivel técnico no universitario. Toma su lugar en la creación de un framework común europeo para la educación y el trabajo donde la necesidad de una herramienta que ayude al entrenamiento de los ciudadanos es más y más necesaria.

Los objetivos de este proyecto son:

- Examinar los problemas relacionados a la definición de competencias y su descripción en los diferentes sistemas europeos a nivel nacional e internacional.
- Selección de los diversos sistemas educativos europeos en el campo de la educación técnica no universitaria a través del análisis de artículos descriptivos y de conceptos

más significativos. En particular, se analiza el Certificate Supplement incluido en el nuevo Europass.

- Comparación de las diferentes herramientas de certificación haciendo uso de las sugerencias incluidas en el documento Europass y en los trabajos de los Puntos Nacionales de Referencia, usando como referencia el esquema ECVET y los niveles ISCED.
- Investigar las características y la cobertura de los documentos sugeridos a nivel de la U.E y al mismo tiempo compararlos con lo que ya existen en los países socios.
- Construir una herramienta común para permitir que el educador complete un documento de certificación de una manera transparente. Tal documento tendrá que considerar la evolución de sugerencias europeas que ya existen y son usadas en los países socios.
- Investigar y experimentar, con las herramientas apropiadas, la comprensión eficaz del certificado emitido, por los usuarios, es decir, los educadores ya que son los encargados de comprobar el correcto funcionamiento del campo educativo y de las competencias adquiridas por la gente cualificada y descritas en los documentos. Tal experimentación debe considerar las herramientas según lo elaborado por los grupos de trabajo técnico sobre la calidad, la transparencia y la transferencia de créditos.
- Promover y dar a conocer las oportunidades ofrecidas por esta herramienta.
- El contexto del proyecto B.E.A.T.R.I.C es el de la educación técnica no universitaria en cooperación con la universidad con el propósito de ofrecer a sus usuarios competencias profesionales que los ayude a incorporarse en el mercado de trabajo. Los resultados previstos del proyecto necesitan ser probados en un campo específico de actividad. Los países socios han elegido como campo de prueba el área turística, un campo fuertemente integrado, caracterizado por una transversalidad común a otros campos.

3.7 European Credits Transfert System for Vocational education and Training (ECVET)

Los Programas educacionales del sector técnico se basan en VET [4] “*Vocational Education Training*”. Estos programas VET son sistemas nacionales de educación diseñados para dar habilidades técnicas a trabajadores y estudiantes que se desarrollarán en sectores particulares del mercado laboral europeo, como plomería, hotelería, ventas, etc.

Cada país confecciona sus programas según las necesidades del mercado laboral, con el objetivo de formar profesionales que se adapten y puedan desarrollarse sin problemas en

un ambiente profesional, por lo mismo es muy importante tener dentro de estos programas educativos tanto horas teóricas en salas de clases, como trabajo práctico en empresas.

Para lograr una comunicación entre las diferentes instituciones educacionales que entregan una formación técnica tanto a nivel nacional como internacional se diseñó el sistema ECVET, que es el “*European Credits Transfert System for Vocational Education and Training*” [6], es decir, “*Sistema Europeo de Transferencia de Créditos para la Educación Vocacional y Entrenamiento*”.

El sistema ECVET fue diseñado con el objetivo de permitir el diálogo e intercambio entre los diferentes sistemas VET, los cuales son muy diferentes en términos de opiniones y organización pedagógica.

ECVET contribuye al aprendizaje de larga vida, facilitando la movilidad de estudiantes dentro de sus trayectorias VET individuales y entre los diferentes sistemas educacionales que proporcionan estos programas, además anima y promueve la movilidad individual geográfica y profesional. Incluye todos los procesos formales, no-formales e informales de entrenamiento, educación y aprendizaje. Por lo tanto, debe apoyar la posibilidad de empleo y desarrollo personal.

Por otro lado requiere y promueve la transparencia de las cualificaciones (definiendo los objetivos de un programa VET, expresados alternadamente en términos de conocimiento, habilidades y competencias), de procedimientos (los procedimientos de carga o de reconocimiento), de los procesos de aprendizaje (formal, no-formal e informal) y de los caminos de aprendizaje (organización de los sistemas VET). Se basa en el establecimiento de la confianza mutua entre las autoridades VET, los establecimientos educacionales en términos de la carga de conocimiento, las habilidades y las competencias adquiridas.

3.7.1 Principios de ECVET

- Permitir a un estudiante moverse desde un sistema nacional formal VET a otro; para tener acceso a un sistema formal VET desde un contexto de aprendizaje no formal; para transferir el aprendizaje resultante entre tales sistemas y el contexto de aprendizaje.
- Acumular y tener un aprendizaje resultante valorado a través de la trayectoria de instrucción VET.
- ECVET está centrado en los estudiantes en el contexto europeo donde la movilidad y la libre circulación de la personas son derechos individuales.

- ECVET es un sistema europeo descentralizado basado en la participación voluntaria de los estados miembros y de los respectivos proveedores VET de acuerdo a la legislación y regulación sobre la carga académica, el reconocimiento, la certificación y la calidad asegurada.
- La implementación de ECVET será apoyada por el Framework común de niveles de referencia.

3.7.2 Framework de Referencia para ECVET

ECVET descansa sobre tres reglas fundamentales, que son los pilares de su efectiva implementación.

Los objetivos de una trayectoria de aprendizaje, de un programa de entrenamiento o de elementos de una cualificación son expresados en términos del conocimiento, habilidades y competencias que se adquirirán y dominarán en un nivel de referencia dado. Ellas se acuerdan formalmente, se juntan y organizan en unidades.

El framework de referencia para ECVET se basa en el acuerdo definido a nivel europeo que un número máximo de créditos es asignado a un conjunto de unidades, correspondiente a una trayectoria de aprendizaje, a un programa VET o a una cualificación completa. Este acuerdo permite asignar a cada unidad (o al conjunto de unidades) un número de puntos del crédito según el peso relativo de cada unidad. El valor de transferencia o intercambio de cada unidad se puede definir en términos de los puntos de crédito.

Los puntos de crédito se asignan a las unidades en base de los resultados del aprendizaje. El valor de una unidad en términos de los puntos de crédito corresponde a una cierta combinación del conocimiento y de las habilidades de competencias. Los resultados del aprendizaje se consideran en relación a niveles de referencia, framework de cualificación y segmentos del mercado laboral (perfiles ocupacionales, familias vocacionales) y puede ser vinculado a referencias existentes nacionales o regionales.

La movilidad individual en un programa de intercambio ocurre entre dos (o más) proveedores VET. Ellos junto con el estudiante, el principiante o el aprendiz interesado, llegan a un acuerdo sobre el contenido y las principales características del programa de movilidad en términos del conocimiento, de las habilidades y de las competencias que serán adquiridas por el estudiante durante su estancia en el exterior. Este acuerdo se coloca en un “*Memorandum of Understanding (MoU)*”.

Un “*Memorandum of Understanding (MoU)*” son acuerdos entre las instituciones sobre los objetivos en términos de conocimientos, habilidades, competencias, duración,

responsabilidades, derechos y obligaciones dentro de los programas de intercambio de movilidad estudiantil. MoU da información en su anexo sobre las estructuras respectivas de enseñanza y las actividades de aprendizaje (de un punto de vista pedagógico y organizacional así como la información sobre la carga de trabajo teórica).

En cualquier contexto nacional o regional los programas de educación o entrenamiento se pueden dividir en unidades o conjunto de unidades. Una unidad de enseñanza se define dentro de ECVET como una parte elemental (o más pequeña) de un plan de estudios o del programa de educación y entrenamiento. Cada unidad corresponde a una combinación específica de conocimientos, habilidades y competencias; que pueden ser de diferente tamaño según los sistemas nacionales VET.

Una lista similar de conocimientos, habilidades y competencias se puede adquirir en cada sistema nacional VET a pesar de las diferencias en las trayectorias de aprendizaje en términos de carga de trabajo teórica, acuerdos sobre el aprendizaje y los métodos de enseñanza. Acordar equivalencias entre las unidades de aprendizaje presentes en los diversos sistemas nacionales, en términos del conocimiento, de las habilidades y de las competencias, significa que es necesario acordar valores iguales para el aprendizaje, los métodos y el tiempo invertido en las actividades de enseñanza.

El valor de las unidades de enseñanza en términos de créditos se basa sobre el acuerdo a nivel europeo que un programa de educación o entrenamiento completo es igual a 180 puntos del crédito. Este acuerdo permite evaluar el valor de intercambio de las unidades VET en términos del conocimiento, habilidades y competencias.

Después del periodo de estudios VET en el exterior, los puntos del crédito adquiridos y el conocimiento, las habilidades y las competencias correspondientes serán integrados en el programa nacional y completados con otras unidades de cualificación dependiendo del programa VET.

El sistema de medida de créditos ECVET, es una iniciativa que mide los conocimientos de una manera cuantificable a través de horas académicas, pero no sirve para medir las competencias de un estudiante, es decir, no evalúa los conocimientos realmente adquiridos por el alumno como es el objetivo de este proyecto, por lo tanto, el modelo no hace una evaluación a nivel práctico de las capacidades.

Capítulo 4

Modelo Conceptual de Integración

4.1 Introducción

Para la formalización de certificados se tomaron como caso de estudio tres países que participan en el proyecto B.E.A.T.R.I.C. Los países escogidos fueron Irlanda, Italia y Holanda. La razón de esta decisión se debe principalmente a que se encontraba a disposición la cantidad suficiente de documentos que daban una información completa del proceso de certificación dentro del framework educativo de cada país, lo que permitió poder realizar cada modelo conceptual lo más consistente posible, debido que éstos serán utilizados para la posterior implementación de la base de datos que sustentará la aplicación ilustrativa.

Al formalizar los certificados de estos países se pudo cuantificar la complejidad de los modelos y así poder encontrar de forma más certera el método de integración que se adecue de mejor forma al problema en estudio que es poder integrar modelos semánticamente distintos.

Los modelos de formalización se desarrollaron en idioma inglés porque la documentación de entrada de los documentos formales se encuentra en la lengua original de cada país o son traducidos al inglés, por lo tanto, se escogió este idioma como estándar.

4.2 Lenguaje de Formalización

Como se explicó en el primer capítulo los diferentes certificados de los países participantes del proyecto de la U.E se encuentran en formato de texto, por lo tanto, una primera contribución de este proyecto fue la formalización de los certificados a través de un mecanismo de modelamiento dentro de los que se barajaban varias opciones como modelos de entidad relación, OWL, UML, redes semánticas, entre otros.

Finalmente se escogió Redes Semánticas como un mecanismo de formalización debido que las personas que utilizarán estos modelos no pertenecen al área informática sino que son personas vinculadas al área educacional, por lo tanto, estos modelos son más fáciles de entender por personas que no tienen una formación en informática.

A continuación se dará una pequeña descripción con respeto al mecanismo de formalización elegido.

4.2.1 Descripción de Redes Semánticas

Una red o una red semántica [28] es una notación gráfica para representar conocimiento en modelos de nodos y arcos interconectados. La implementación computacional de redes semánticas fue primero desarrollada para la inteligencia artificial, pero versiones anteriores se han utilizado por mucho tiempo en filosofía, psicología, y lingüística.

Lo que es común a todas las redes semánticas es una representación gráfica que se puede utilizar para representar conocimiento o para apoyar sistemas automatizados para el razonamiento sobre el conocimiento. Algunas versiones son altamente informales, pero otras son sistemas formalmente definidos por la lógica. Las siguientes son las seis clases de redes semánticas más comunes.

- Las *Redes de Definición*, enfatizan el subtipo o una relación entre un tipo de concepto y un subtipo nuevamente definido. La red resultante, también llamada una generalización, apoya la regla de la herencia para copiar propiedades definidas para un súper tipo a todos sus subtipos. Puesto que las definiciones son verdaderas por definición, la información en estas redes frecuentemente se asume como necesariamente verdad.
- Las *Redes de Afirmación* se diseñan para afirmar propósitos. Difieren de las redes definicionales, la información en una red de afirmación se asume que será circunstancialmente verdadera, a menos que esté explícitamente marcada como un operador modal. Algunas redes de afirmación se han propuesto como modelos de estructuras conceptuales inferiores a la semántica del lenguaje natural.
- Las *Redes de Implicación*, utilizan la implicación como la relación primaria para conectar nodos. Pueden ser utilizadas para representar modelos de verdad, de causalidad, o de inferencias.
- Las *Redes de Ejecución*, incluyen algunos mecanismos, tales como marcar pasos o unir procedimientos, que pueden realizar inferencias, pasar mensajes, o búsquedas para modelos y asociaciones.
- Las *Redes de Aprendizaje*, se construyen o extienden sus representaciones adquiriendo conocimiento de ejemplos. El nuevo conocimiento puede cambiar la red antigua agregando y suprimiendo nodos y arcos o modificando los valores numéricos, llamados pesos, asociados a los nodos y a los arcos.

- Las *Redes Híbridas* combinan dos o más de las técnicas anteriores, en una sola red o en redes separadas.

Algunas de las redes se han diseñado explícitamente para implementar hipótesis sobre mecanismos cognoscitivos de humanos, mientras que otras han sido diseñadas principalmente para formalizar conceptos del mundo real explotando sus características informativas.

Las notaciones de redes y las notaciones lineales son capaces de expresar información equivalente, pero ciertos mecanismos de representación se integran mejor de una forma que de otra. Puesto que los límites lineales son vagos, por lo tanto, es imposible dar condiciones necesarias y suficientes que incluyan todas las redes semánticas mientras se excluyen otros sistemas que no son llamados generalmente redes semánticas.

4.2.2 Notación Formal Basada Sobre Redes Semánticas

Una red semántica [28] es un diagrama que representa las conexiones (relaciones) entre los conceptos en algunos dominios específicos de conocimiento. Tales relaciones existen en la mente humana y los diagramas son justo una manera tentativa de representar lo que pensamos, el conocimiento almacenado, los conceptos y las relaciones entre ellas.

Una red semántica es un diagrama orientado que consiste de una serie de nodos que están conectados unos a otros a través de arcos dirigidos y etiquetados. Además, se pueden definir restricciones sobre estos nodos y arcos. En un nivel abstracto, un modelo de red semántica consiste de cuatro componentes:

- Un conjunto de nodos que representan objetos del mundo real.
- Un conjunto de arcos dirigidos que representan las relaciones semánticas entre los nodos.
- Un conjunto de etiquetas que denotan los diferentes tipos de relaciones semánticas.
- Un conjunto de restricciones definidas sobre los nodos y arcos.

Para las redes se establecerán dos categorías de nodos, nodos complejos y nodos básicos.

Los nodos básicos no tendrán ningún arco que salga de ellos. Es decir, son los nodos hoja en el diagrama de red semántica.

Los nodos complejos son los nodos internos en el diagrama de la red. Cada nodo complejo tiene uno o más arcos dirigidos etiquetados que salen de él, cada uno asociado con una etiqueta, que va a otro nodo. La figura 4.1 da un ejemplo simple de una parte de una red semántica diseñada según la metodología propuesta para el sistema de certificación irlandés FETAC, donde se denotan los nodos complejos usando una caja y los nodos básicos usando un punto.

Un nodo complejo en este ejemplo es “*FETAC Advanced Certificate*”, que tiene unos o más arcos salientes. Los nodos básicos son “*Title Certificate*”, “*Date of Release*”, “*Chair*”, “*Executive Chief*”, “*Award Type*” y “*Purpose*”.

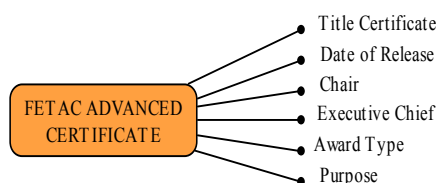


Figura 4.1. Representación de un Nodo Complejo y Varios Nodos Básicos.

Cada arco en el modelo vincula dos nodos, representando su relación semántica. Esta relación binaria expresa la conexión estática del inter-objeto, y por lo tanto captura los aspectos estructurales de los objetos del mundo real. Hay tres maneras de acoplar los nodos dentro de una red que son:

La *generalización* que es una relación entre un concepto general y un concepto más específico. A veces la relación es llamada “*es*” o “*es una clase*”, se utiliza frecuentemente para mostrar herencia entre los objetos.

La *asociación* es una relación estructural, que especifica qué objetos de una clase están conectados con objetos de otra. Dada una asociación que conecta dos objetos, se puede navegar de uno a otro, y viceversa. Una asociación entre dos objetos representa una relación estructural entre los pares, significa que ambos conceptos están al mismo nivel, y que ninguno es más importante que otro.

La *agregación* es una relación compuesta, en la cual un objeto compuesto ("completo") consiste de algunos objetos componentes ("piezas"). Esta clase de relación representa "parte" de la relación. La agregación se trata a menudo como clase de asociación. Pero los objetos en la relación de agregación no son conceptos que están en el mismo nivel. Esta característica distingue la relación de agregación de la relación general de asociación.

4.2.2.1 Generalización

Un concepto muy importante en el diseño orientado a objeto está estrictamente relacionado al concepto de generalización [28], es decir, la herencia. Se refiere a la capacidad de un concepto (o nodo hijo) de heredar las funciones idénticas de otro concepto (o súper nodo).

Para modelar la herencia en una red semántica, se dibuja una línea desde el concepto hijo (el concepto que hereda el comportamiento) con una flecha que señala al súper concepto. Considere el siguiente ejemplo del esquema italiano de certificación IFTS [28] mostrado en la figura 4.2. Aquí, las competencias para cada área se articulan en unidades que se han construido alrededor de núcleos conceptuales fundamentales. Las unidades se aceptan como elementos funcionales para la determinación de las competencias que cada persona debe poseer. Así, las unidades básicas y las unidades cruzadas se pueden identificar como agregación de capacidades básicas y cruzadas. Estas unidades (con sus capacidades asociadas) complementan las Unidades y competencias técnico-sectoriales específicas, asociados a una figura vocacional particular.

La dirección de la flecha es importante; ya que la relación se puede leer solamente en esa dirección. Por lo tanto, sería incorrecto decir que "una unidad es una clase de unidad básica". Siguiendo el camino de las flechas los subtipos heredan las características del súper tipo, así por ejemplo, podemos decir que las unidades cruzadas, básicas y técnica-sectoriales tienen un título, característica heredada de la súper clase unidad

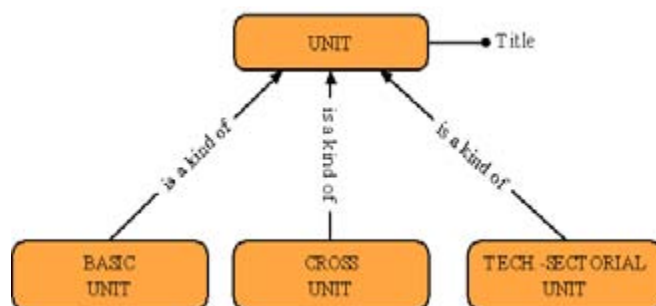


Figura 4.2. Representación Formal de la Generalización.

4.2.2.2 Asociación

Cuando se modela un sistema, ciertos conceptos serán relacionados unos a otros, y estas relaciones necesitan ser modeladas con claridad. Hay varios tipos de asociaciones. Se Consideraran las asociaciones unidireccionales para modelar la mayoría de las relaciones dentro de los framework de certificación considerados. Una asociación unidireccional [28] demuestra que dos conceptos están relacionados y expresa la dirección de navegabilidad de la relación.

La figura 4.3 muestra un ejemplo del modelo italiano de certificación, donde hay una relación unidireccional entre los nodos “*Final IFTS Certificate*” y “*Other Ed./Train. Institution*”. Una asociación unidireccional se dibuja como una línea con una punta de flecha que señala al concepto asociado. La asociación unidireccional incluye el nombre y el valor de la multiplicidad. El nombre detalla el significado de la relación y la dirección la indica la punta de la flecha. En el ejemplo considerado, el significado de la relación es el siguiente: la trayectoria que conduce a un “*Final IFTS Certificate*” implica al concepto de “*Other Ed./Train. Institution*”.

Diversas restricciones se pueden especificar en la red semántica para realzar su capacidad de capturar los significados semánticos del mundo real. Estas restricciones se pueden definir sobre nodos y arcos. Estas explican varios requerimientos tales como dominio, cardinalidad, orden, homogeneidad, heterogeneidad y unicidad. Básicamente, la restricción más utilizada es la cardinalidad.

La cardinalidad especifica el número de instancias de un nodo que se pueden relacionar a una sola instancia del otro nodo asociado; o el número de instancias que participan como una “parte” de otra instancia completa. Típicamente las restricciones de cardinalidad incluyen “*uno a uno*” [1..1], “*uno ó muchos*” [1..N], “*cero ó uno*” [0..1], y “*cero ó muchos*” [0..N].

Específicamente, según la multiplicidad de la asociación se puede decir que la trayectoria hacia la asignación de un “*Final IFTS Certificate*” implica una o más “*Other Ed./Train. Institution*”. Leyendo la relación en dirección opuesta, se puede decir que una o más “*Other Ed./Train. Institution*” son necesarias para el desarrollo de un “*Final IFTS Certificate*”. Por lo tanto, la dirección se asume como bidireccional. Una asociación bidireccional es indicada por una línea entre los dos conceptos. El nombre de la asociación se pone sobre la línea y se insertan las multiplicidades. Esta clase de asociación no se utilizará en los diagramas semánticos bajo consideración en este proyecto.

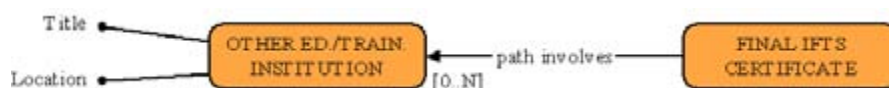


Figura 4.3. Representación formal de las restricciones de cardinalidad en una red semántica.

4.2.2.3 Agregación

La agregación [28] es un tipo especial de relación usado para modelar “El todo es una relación de sus partes”. Una asociación con una relación de agregación indica que un concepto es un concepto subordinado (o una parte) de otro concepto. En una relación de agregación, la clase hijo puede sobrevivir a su concepto padre. Para representar una

relación de agregación, se dibuja una línea desde el concepto padre a la clase subordinada y se utiliza un nombre conveniente para la relación. Figura 4.4 muestra un ejemplo de una relación de agregación entre los conceptos “*Education and Training*”, “*School -/training centre based*”, “*Workplace based*” y “*Accredited prior learning*” dentro del framework Certificate Supplement.

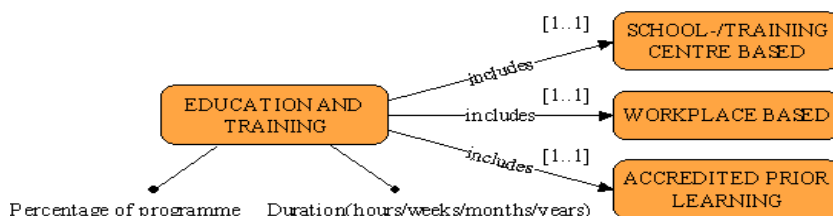


Figura 4.4. Representación formal de la agregación dentro de una red semántica.

4.3 Certificate Supplement

El *Certificate Supplement* (C.S) no es un sustituto para el certificado original. Es complementario a éste, ya que contiene una descripción detallada de las habilidades y capacidades adquiridas por el poseedor del certificado, además de describir el rango ocupacional al cual puede acceder la persona.

El C.S. es elaborado por autoridades que entregan una información detallada de las habilidades y competencias adquiridas por el poseedor del certificado, describen los sectores ocupacionales a los cuales éste tiene acceso, junto con entregar información sobre las instituciones que se encargan de la certificación, el nivel del certificado, las diversas formas de adquirir uno, los requisitos de entrada y las oportunidades de acceder a un siguiente nivel educacional y los puntos nacionales de referencia. Este certificado además puede ser traducido a otro idioma por petición del interesado.

Debido a la diversidad de la información que se encuentra en el certificado, a su dimensión multilingüe, multinacional, se deben respetar un conjunto de normas entregadas por la Unión Europea al momento de completar el certificado para asegurar la consistencia del contenido.

Al tener una representación formal del documento, se tiene el meta-modelo necesario que servirá como nexo para poder realizar el mapping entre los modelos de los diferentes países. La figura 4.5 y 4.6 muestra el documento *Certificate Supplement*.



Europass certificate supplement ^(*)



Ireland

1. Title of the certificate ⁽¹⁾
National Craft certificate: Electrician
⁽¹⁾ In the original language
2. Translated title of the certificate ⁽¹⁾
⁽¹⁾ If applicable. This translation has no legal status.
3. Profile of skills and competences
<ul style="list-style-type: none"> - Plan the job and estimate the amount of material required, from supplied drawings; - Interpret schematic diagrams and flow charts; - Utilise electrical test equipment; - Install and test wiring systems for lighting and power distribution; - Install and test switch gear and distribution boards; - Locate and rectify faults in wiring systems; - Locate and rectify faults in electrical equipment; - Install, test, commission and maintain lighting fittings and controls; - Install, test, commission and maintain electrical heating devises and controls; - Install, test, commission and maintain motive power devises; - Complete job related documentation.
4. Range of occupations accessible to the holder of the certificate ⁽¹⁾
Installation Electrician; Maintenance Electrician; Data Communications Systems Installer; Alarm Systems Installer.
⁽¹⁾ If applicable
<p>^(*) Explanatory note</p> <p>This document is designed to provide additional information about the specified certificate and does not have any legal status in itself. The format of the description is based on the following texts: Council Resolution 93/C 49/01 of 3 December 1992 on the transparency of qualifications, Council Resolution 96/C 224/04 of 15 July 1996 on the transparency of vocational training certificates, and Recommendation 2001/613/EC of the European Parliament and of the Council of 10 July 2001 on mobility within the Community for students, persons undergoing training, volunteers, teachers and trainers.</p> <p>More information available at: http://europass.cedefop.eu.int</p> <p>© European Communities 2002</p>

Figura 4.5. Formato del Certificate Supplement (hoja 1).

5. Official basis of the certificate		
Name and status of the body awarding the certificate FAS - Training and Employment Authority; - Department of Education and Science. <i>FAS - Training and Employment Authority, was established in 1985 under the Labour Services Act 1987 to provide a wide range of services to the labour market in Ireland. FAS is the statutory authority for apprenticeship in Ireland and has statutory powers of certification.</i>		Name and status of the national/regional authority providing accreditation/recognition of the certificate
Level of the certificate (national or international) Craft	Grading scale / Pass requirements Pass: - Practical: all essential skills passed in all phases, including on the job assessment, Theory: 70 %. Merit: Practical: all essential skills plus desirable skills passed in all phases, - Theory: 85 %	
Access to next level of education/training Access to third level Legal basis Labour Services Act 1987, Industrial Training Act 1987.	International agreements	
6. Officially recognised ways of acquiring the certificate		
Description of vocational education and training received	Percentage of total programme (%)	Duration (hours/weeks/months/years)
Training centre-based	10%	20 weeks
Workplace based	10%	20 weeks
Accredited prior learning	80%	3.2 years
Total duration of the education/ training leading to the certificate		4 years
entry requirements Minimum of grade 'B' in two (2) subjects in the Department of Education's Junior Certificate or its approved equivalent.		
Additional information More information (including a description of the national qualifications system) available at: http://www.fas.ie		
More information (including a description of the national qualifications system) available at: www.waw		
National reference point http://www.nqai.ie/nqai/pubs/indexpage.htm		

Figura 4.6. Formato del Certificate Supplement (hoja 2).

4.3.1 Descripción de la Red Semántica del Certificate Supplement

- **Certificate Supplement:** Es el certificado mismo que ha sido modelado como un nodo complejo. La razón de esta elección se debe que este concepto actúa como un contenedor de un gran número de conceptos que se detallarán más adelante. Los nodos hojas que contiene este nodo son el título del certificado, que corresponde al título profesional del poseedor de éste en lenguaje original, lenguaje según el código ISO que describe el idioma en que se encuentra el título del certificado, éste es un campo fijo, traducción del título este campo se llena sólo si el título debe ser traducido, siguiendo ciertas pautas de traducción, ésta debe ser lo más literal con respecto al título original, nombre del país que entrega el certificado, que corresponde al país en el cual el poseedor del certificado estudió.
- **Profile of Skills and Competences:** Este nodo complejo da una descripción precisa de las competencias obtenidas por el poseedor del certificado durante su periodo académico. Esta lista por lo general comienza con “*Un poseedor de este certificado es capaz de*” y se incluye un conjunto de 5 a 15 competencias utilizando verbos que las describan, por ejemplo “*Planear el trabajo y estimar la cantidad de material requerido para realizar el abastecimiento*”. Se entrega un glosario de referencia en múltiples idiomas sobre la terminología para poder completar y traducir el certificado. Las palabras que mencionan estas habilidades y competencias deben tener la siguiente estructura: verbos + objetos + complementos, es decir, la descripción debe ser concisa y evitar adverbios que describan actitudes así como cualquier elemento que no proporcione información esencial y por último éstas no deben expresar juicios valóricos. Para dar una formalización consistente al documento C.S, se decidió mantener el concepto de *Profile of Skills and Competences* y agregar un concepto adicional denominado *Skills and Competences*. Por lo tanto el *Profile of Skills and Competence*, se ha modelado como un nodo complejo que no contendrá atributos, sin embargo se agregará una relación que vincule los dos nodos antes mencionados, el nombre de esta relación será “*compuesta por*” y la restricción de cardinalidad será [1..N], es decir, el *Profile of Skills and Competences* estará compuesto por al menos 1 o muchas *Skills and Competences*.
- **Skill/Competence:** Como ya se mencionó el *Profile of Skills and Competences* se caracteriza por un conjunto de habilidades y competencias, cada una descrita según las pautas entregadas por la Unión Europea. Una estrategia de

formalización fue crear este nodo simple y vincularlo directamente al nodo mencionado en el punto anterior.

- **Range of Occupations Accessible to the Holder of the Certificate:** Este campo dentro del documento es opcional, ya que describe el conjunto de campos ocupacionales a los cuales puede acceder el poseedor del certificado. Para el modelo este es un nodo complejo que contiene un atributo que describe cada una de los campos ocupacionales a los que puede ingresar el poseedor del certificado. Se estableció una relación que vincula este nodo al nodo complejo *Certificate Supplement* cuya restricción de cardinalidad es [0..N].
- **Official Basis of the Certificate:** En este nodo complejo se debe describir el nombre de la institución que entrega el certificado, el nombre de la autoridad nacional y regional que da el reconocimiento de acreditación, el nivel del certificado dentro del framework nacional, la escala de evaluación, el siguiente nivel educacional al que puede acceder el poseedor del certificado, acuerdos internacionales y la base legal.
- **Body Awarding the Certificate:** El nombre completo de la institución y el tipo (pública, regional privada, nacional, etc.) así como la dirección, el número telefónico, el fax de existir, la dirección Web y dirección de correo electrónico deben ser especificados.
Este es un nodo complejo, está directamente relacionado al nodo *Official Basis of the Certificate* a través de una relación con restricción de cardinalidad [1..1].
- **National/Regional Authority Providing Accreditation/Recognition of the Certificate:** En este nodo complejo se debe especificar si el certificado es acreditado por una autoridad nacional o regional, el nombre completo de ésta y el tipo de institución (pública, regional, privada, nacional, etc.), junto con la dirección, el número telefónico, el número de fax de existir, la dirección Web y la dirección de correo electrónico. Esta autoridad es distinta del awarding body.
Este concepto es opcional, por esto una restricción de cardinalidad mínima [0..1] debió ser usada para la relación que la vincula al concepto de *Official Basis of the Certificate*.
Si es necesario, el periodo de validez del certificado puede ser indicado.
- **Officially Recognised Ways of Acquiring the Certificate:** Las diferentes maneras de adquirir el certificado deben ser indicadas en el documento. Estas pueden ser por un entrenamiento alternado, continuo, entrenamiento basado en los lugares de trabajo o en colegio. Por lo tanto, los requisitos de acceso a cada

una de estas formas de entrenamiento deben ser especificados, además de cualquier tipo de información adicional y los puntos nacionales de referencia.

Este es un nodo complejo dentro de la red semántica que contiene una serie de campos que describen la información mencionada.

Se debe especificar la duración total de cada uno de los programas académicos seguidos por el poseedor del certificado. Por ello, el concepto de Duración total se ha agregado al modelo como un nodo básico u hoja.

El nodo *Officially Recognized Ways of Acquiring the Certificate* se ha vinculado al nodo *Certificate Supplement* a través de una restricción de cardinalidad [1..1].

El significado de esta relación es el siguiente el *Certificate Supplement* especifica la *Officially Recognized Ways of Acquiring the Certificate*.

En particular el nodo *Vocational Education/Training* se ha introducido en el modelo. Esta es una generalización que representa directamente a la tabla presente en el documento C.S. Entonces, para cada una de las filas de esta tabla se han insertado dentro del modelo tres conceptos llamados *School/Training Centre-Based*, *Workplace-Based* y *Accredited Prior Learning*. El concepto más general *Vocational Education/Training* se ha vinculado al nodo *Officially Recognized Ways of Acquiring the Certificate* a través de una restricción de cardinalidad [1..N].

Más adelante, también los conceptos requisitos de entrada, información adicional y los puntos nacionales de referencia serán modelados como nodos complejos. Estos conceptos se vincularán al nodo *Officially Recognized Ways of Acquiring the Certificate* a través de una restricción de cardinalidad [1..N].

- **Vocational Education/Training:** Este nodo a nivel abstracto representa una de las filas dentro del documento C.S y define las experiencias educacionales y de entrenamiento del poseedor del certificado. Como ya se mencionó este nodo complejo se vincula por una relación [1..N] al nodo *Officially Recognized Ways of Acquiring the Certificate*, esto significa que una o más experiencias vocacionales de entrenamiento pueden ser incluidas dentro del C.S.
- **School-/Training Centre-Based:** Este nodo representa una fila dentro del C.S la cual identifica la situación educacional y de entrenamiento del poseedor del certificado. Se representa como un nodo complejo ya que representa una especificación del nodo *Vocational Education/Training*, de él hereda sus características. La relación que vincula este nodo con el concepto *Vocational Education/Training* es [1..1].
- **Workplace-Based:** Este nodo representa una fila dentro del C.S la cual identifica la situación educacional y de entrenamiento del poseedor del

certificado. Se representa como un nodo complejo ya que representa una especificación del nodo *Vocational Education/Training*, de él hereda sus características. La relación que vincula este nodo con el concepto *Vocational Education/Training* es [1..1].

- **Accredited Prior Learning:** Este nodo representa una fila dentro del C.S la cual identifica la situación educacional y de entrenamiento del poseedor del certificado. Se representa como un nodo complejo ya que representa una especificación del nodo *Vocational Education/Training*, de él hereda sus características. La relación que vincula este nodo con el concepto *Vocational Education/Training* es [1..1].
- **Entry/Access Requirement:** Este es un concepto opcional utilizado para especificar los requisitos educacionales para cada nivel, como por ejemplo el nivel educacional necesario, el modo de selección, etc. Por lo tanto, varios requisitos de entrada pueden ser especificados. Entonces este nodo complejo, que posee un nodo básico llamado descripción se relaciona al nodo *Officially Recognized Ways of Acquiring the Certificate* por una restricción de cardinalidad [0.N].
- **Additional Information:** Este nodo se utiliza para proporcionar una breve información sobre la institución que se encargó de educar al poseedor del certificado, por ejemplo el tipo de entrenamiento práctico, el programa o estructura del sistema educacional vocacional, etc. Si es relevante también describe la dirección del sitio Web que entrega información sobre el sistema de cualificación nacional. Por lo tanto, se puede agregar mucha información adicional. Este nodo complejo se vincula al nodo *Officially Recognized Ways of Acquiring the Certificate* por una restricción de cardinalidad [0.N].
- **National Reference Point:** De acuerdo a las guías para completar el Certificate Supplement, el nombre y dirección incluida la dirección web del punto nacional de referencia deben ser incluidas dentro del certificado. Por lo tanto, este concepto contiene más de un subconcepto que debieron ser formalizados en el modelo. Entonces este nodo está relacionado al nodo *Officially Recognized Ways of Acquiring the Certificate* por una restricción de cardinalidad [1..1].

La figura 4.7 muestra la red semántica del Certificate Supplement.

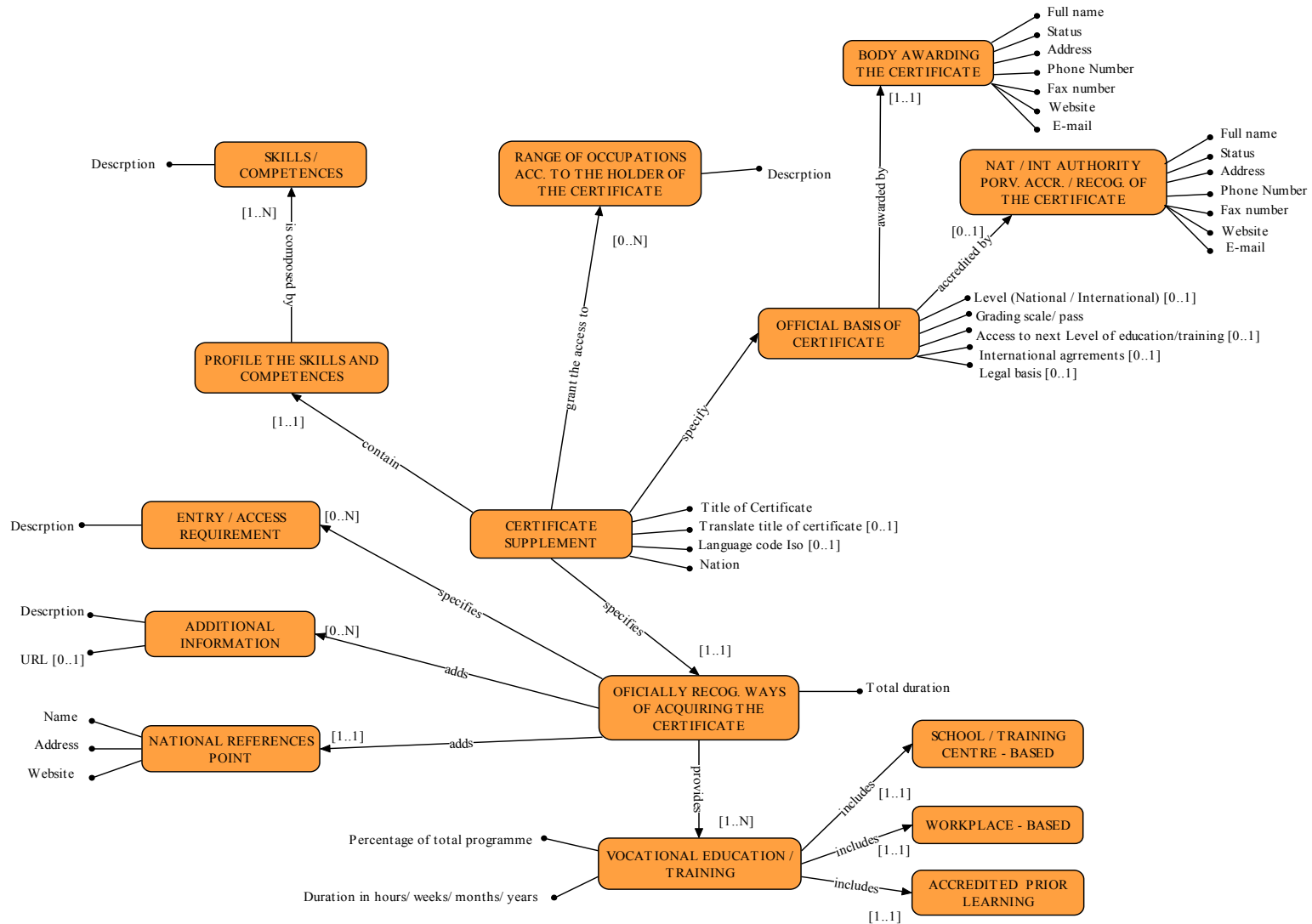


Figura 4.7. Red Semántica del Certificate Supplement.

4.4 Italia

4.4.1 Introducción

A continuación se describirá el proceso de certificación de Italia, uno de los países que se seleccionó como caso de estudio para poder tener una visión más amplia en cuanto a lo que es la certificación en Europa y para posteriormente comprobar la utilidad de Information Retrieval en la comparación semántica de competencias con el fin de poder determinar si dos programas académicos de distintos países pueden ser homologados.

Se seleccionó Italia como caso de estudio debido a la cantidad de información disponible con lo que se pudo obtener un modelo semántico consistente, por lo tanto, dentro de este punto también se describirá la red semántica generada a partir de certificado IFTS.

4.4.2 Certificación en Italia

El sistema IFTS, *Higher Technical Education and Training*, es un vía de entrenamiento de educación técnica no universitaria [10]. Su objetivo es acercarse y encontrar respuestas a las necesidades del mercado laboral de una manera interactiva. Los cursos IFTS [8] se planean en conjunto entre el ministerio de educación, las universidades, el ministerio del trabajo regional, agencias locales y comités nacionales de planeación que definen guías, procedimientos y estándares nacionales mínimos.

Los cursos se manejan y administran de una manera integrada de al menos cuatro tipos de instituciones educacionales: colegios, institutos, universidades y empresas privadas o públicas. La duración de cada uno es de 2 a 4 semestres.

Al final de cada curso se realiza una prueba de evaluación, donde se entrega el “*Certification of Higher Technical Education*”, esta certificación corresponde al cuarto nivel dentro del *European Qualification Framework* [9].

4.4.3 Descripción del Certificado IFTS

Los indicadores incluidos dentro del modelo del certificado tienen el objetivo de dar información esencial con respecto a la experiencia formativa, conocimientos y competencias adquiridas así como las principales características de la trayectoria educacional del alumno de una manera resumida.

El certificado que se muestra a continuación en las figuras 4.8 y 4.9 corresponde a un certificado IFTS y fue formalizado a través de una red semántica, ya que es un

requerimiento propio del proyecto B.E.A.T.R.I.C utilizar redes semánticas como mecanismo de modelamiento.

Página 1

(MODELLO DI CERTIFICATO)
 (Legge Regionale, Unione Europea, Ministero Istruzione Università e Ricerca)

CERTIFICATO DI SPECIALIZZAZIONE TECNICA SUPERIORE

1.1);

RILASCIATO A CONCLUSIONE DEL PERCORSO DI ISTRUZIONE E FORMAZIONE TECNICA SUPERIORE ISTITUITO CON DELIBERA REGIONALE N.

Livello Europeo IV

CONFERITO (2)

A

1.2);

Soggetti attuatori del corso

Denominazione	Codice
.....
.....

L'ASSESSORE
(o suo delegato)

IL PRESIDENTE DELLA
COMMISSIONE REGIONALE IFTS

Registrazione Data:

Il presente corso è stato realizzato in conformità con le norme stabilite dal regolamento (CE) n. 2609/2002 della Commissione Europea e del regolamento (CE) n. 2609/2002 del Parlamento Europeo. Nel caso di cui al presente articolo non è previsto nel campo di applicazione la data di scadenza per il rilascio o la concessione del presente corso di formazione o l'iscrizione del corso stesso. Il presente corso è stato realizzato nel 2002/2003. L'assessorato Regionale Regionale.....

Il presente corso è stato realizzato in conformità con le norme stabilite dal regolamento (CE) n. 2609/2002 della Commissione Europea e del regolamento (CE) n. 2609/2002 del Parlamento Europeo. Nel caso di cui al presente articolo non è previsto nel campo di applicazione la data di scadenza per il rilascio o la concessione del presente corso di formazione o l'iscrizione del corso stesso. Il presente corso è stato realizzato nel 2002/2003. L'assessorato Regionale Regionale.....

Página 2

4. PERCORSO FORMATIVO

4.1) DURATA: VCEI CFE

4.2) UNITÀ FORMATIVE

Unità formative	Durata (anni)	Quota di apprendimento	Modalità di valutazione
.....
.....
.....

Figura 4.8. Formato del certificado IFTS (página 1).

Página 4

4. INFORMAZIONI

Finalità (scienze, Lettere, Giurisprudenza, Scienze Politiche)	Durata (anni)	Settore e competenza di area (la)	Dimensione (numero di studenti)	Indirizzo (sezione o specializzazione)
.....

5. MODALITÀ DI VALUTAZIONE FINALE

5.1. Esame di Stato

.....
.....

5.2. Esame di Laurea

.....
.....

6. CREDITI FORMATIVI

.....
.....

Pagina 5

7. ANNOTAZIONI INTEGRATIVE/ALTRE INFORMAZIONI

.....
.....
.....

Figura 4.9. Formato del certificado IFTS (página 2).

4.4.4 Descripción de la Red Semántica del Certificado IFTS

La red semántica del certificado IFTS se diseñó a través del certificado de las figuras 4.8 y 4.9 y un conjunto de documentos entregados que describían el sistema de certificación italiano a nivel técnico no universitario. A continuación se describirán cada uno de los nodos complejos que conforman la red semántica para el certificado italiano.

- **Final Certification:** Al final de cada uno de los cursos del IFTS, las regiones entregan un certificado de acuerdo al modelo entregado junto con las guías para su llenado. La certificación final representa un elemento conceptual base dentro

del sistemas de certificación IFTS, además muestra las relaciones con otros elementos conceptuales e ítems descriptivos dentro del sistema. Este nodo incluye información sobre los siguientes campos: título del certificado, fecha y lugar de la entrega, nombre de la región que realiza la entrega del certificado, el nivel de éste, el nombre del supervisor regional y del presidente de la comisión a cargo de la examinación final. Por lo tanto el “*Final Certification*” ha sido modelado como un nodo complejo dentro de la red semántica.

- **Receiver:** Para evitar cualquier tipo de discriminación la información incluida dentro de este nodo complejo se debe limitar sólo a criterios estrictamente necesarios para su identificación. Por lo tanto los nodos hojas que contendrá serán el nombre, la fecha y el lugar de nacimiento del destinatario. Por lo tanto, fue modelado como un nodo complejo.
Este concepto está estrictamente relacionado al nodo “*Final Certification*”. Esta relación está caracterizada por una relación de cardinalidad [1..1], es decir, un certificado se entrega a un único destinatario.
- **Structure Handling the Path:** El nombre de la institución donde se realizó el proceso formativo del estudiante se considera un elemento esencial, especialmente para las empresas. Ya que no sólo muestra el proceso formativo del alumno sino que también da una correspondencia directa entre la adquisición del certificado y las capacidades profesionales del alumno que muestra a los empresarios el nivel de calidad de la institución formativa.
Dentro de este nodo se detallan los siguientes campos, el nombre de la institución y la dirección. Por lo tanto es un nodo complejo que está vinculado al nodo “*Final Certification*” donde la relación tiene una restricción de cardinalidad [1..1], es decir, un “*Final Certification*” puede ser manejada por sólo una “*Structure Handling the Path*”.
- **Additional Educational Structure:** Este nodo agrupa información sobre cualquier otra institución que por algún acuerdo de cooperación haya participado a lo largo del camino educacional del alumno, los campos que se describen dentro de este nodo son el nombre y la dirección del establecimiento. Este nodo complejo está relacionado al “*Final Certification*” a través de una relación con una restricción de cardinalidad de [1..N], es decir, en un “*Final Certification*” pueden participar una o muchas “*Additional Educational Structure*”.
- **Vocational Figure of Reference:** En este nodo se detallan las características principales de la figura profesional. Se incluyen los siguientes campos: nombre

de la figura vocacional, clasificación ISTAT, actividades profesionales fundamentales, descripción de cada una de las competencias y habilidades asociadas a la actividad profesional. Este nodo tiene una relación con el “*Final Certification*”, es decir, describe una “*Vocational Figure of Referente*”, la restricción de cardinalidad de esta relación es [1..1].

- **Regional Professional Profile:** Este nodo sólo contiene información cuando se utiliza un perfil profesional regional, como referencia para la confección del plan de estudios del alumno. Por lo tanto este nodo se relaciona al nodo “*Final Certification*” a través de una restricción de cardinalidad [0..1].

- **Achieved Competences:** El concepto competencia asume varias connotaciones y formas dependiendo de los diferentes enfoques y modelos usados para definirlos. Cuando se define el concepto competencia con el objetivo de certificación, el término se usa para identificar el conjunto de recursos (conocimientos, habilidades, etc.) que un alumno debe tener para enfrentar eficazmente el ingreso al mundo laboral, además de dirigir su crecimiento profesional y personal.

La descripción detallada de competencia destaca un aspecto fundamental para el proceso formativo de certificación ya que ayuda a favorecer su transparencia dentro del mercado laboral.

Las competencias expresan relaciones entre un tema y situaciones específicas, por lo tanto, no pueden ser descritas por un análisis exhaustivo de la naturaleza técnica de las actividades laborales ni por la definición de una suma de conocimientos y habilidades de sólo una parte de un determinado tema. En otras palabras se deben definir considerando la clase de recursos que se explotaran y la manera en que se combinaran para enfrentar las diversas situaciones profesionales.

Por todo esto las competencias constituyen un concepto sistemático, un conjunto integrado de recursos mixtos de diferente naturaleza (conocimientos, habilidades, recursos personales, etc.) que se combinan dinámicamente para la aplicación de un programa académico que formará al alumno para actividad laboral.

Las competencias alcanzadas dentro de un curso IFTS se clasifican dentro de tres categorías (básica, cruzadas y competencias técnico-sectoriales específicas) las cuales corresponden a los tres tipos de recursos que combina un programa académico para una efectiva enseñanza dentro del contexto laboral. Por lo tanto, hay tres relaciones dentro de la red semántica que vinculan este nodo con “*Basic competentes*”, “*Cross Competences*” y “*Technical-Sectorial Competentes*”. La restricción de cardinalidad de cada una de estas relaciones es [0..N]. Esto significa que el nodo de certificación permite insertar de cero a

muchas competencias básica, cero o muchas competencias cruzadas y cero o muchas competencias técnico-sectoriales.

- **Basic Competence:** Este nodo describe las competencias técnicas básicas que se necesitan entregar para dar acceso al mercado laboral, así como para el desarrollo de la formación profesional y personal. Reconociendo la contribución de los colegios y universidades en el desarrollo de los programas profesionales, entregando una visión más global con respecto a las competencias básicas. El concepto de competencias básicas se refiere al conjunto de conocimientos que constituyen los requerimientos de entrada a cualquier tipo de educación adicional y la base mínima para ingresar a un trabajo o profesión. Las capacidades básicas en idioma, computación, economía y leyes son algunos ejemplos de este tipo de competencias. Por lo tanto, el desarrollo de un amplio rango de competencias básicas es hoy el objetivo conjunto de colegios e institutos técnicos así como de las universidades, cada uno dentro de su respectivo contexto.
- **Cross Competence:** Este nodo describe aquellas competencias relacionadas a la comunicación, relaciones y solución de problemas que juegan un papel importante dentro de las situaciones laborales y permite transformar los conocimientos en un comportamiento adecuado dentro de cada contexto específico. Es importante destacar como dentro de las competencias cruzadas todos los recursos cognitivos y metodológicos entregados por los colegios y universidades son tomados en consideración.
- **Technical-Sectorial Competence:** Este tipo de competencias describe los conocimientos y técnicas relacionadas a la ejecución de las actividades operativas dentro del proceso laboral, las cuales hacen referencia a distintos contextos dentro del mercado del trabajo.
- **Unit of Competence:** Las competencias se han dividido dentro de tres categorías ya mencionadas, pero éstas pueden ser estructuradas dentro de unidades que permiten evitar una excesiva fragmentación en la descripción de las competencias. La agregación de competencias dentro de unidades ya es muy utilizada en varios países de Europa (Francia, España, Reino Unido). Las unidades de competencias en Italia han sido diseñadas como un conjunto expresivo de las competencias mismas que se pueden reconocer dentro del mercado laboral como un componente de profesionalismo específico y puede ser identificado como el resultado esperado del proceso formativo tanto para las empresas como para las instituciones educacionales. Por lo tanto, las unidades de competencias describen un conjunto más detallado de competencias básicas,

cruzadas y técnico-sectoriales. Por lo tanto, este nodo es una generalización que describe las “*Basic Unit*”, “*Cross Unit*” y “*Technical-Sectorial Unit*”.

- **Basic Unit:** Las unidades básicas de competencias corresponden a una reagrupación de las competencias básicas. Por lo tanto, para una descripción detallada de ellas es necesario analizar la descripción de su competencia asociada. Este nodo está relacionado al nodo “*Basic Competentes*”, cuya restricción de cardinalidad es [1..1], es decir, una “*Basic Competentes*” puede ser agrupada en al menos una y sólo una “*Basic Unit*”.
- **Cross Unit:** Las unidades cruzadas de competencias corresponden a una reagrupación de las competencias cruzadas. Para dar una descripción detallada de ellas es necesario analizar la descripción de su competencia asociada. Este nodo está relacionado al nodo “*Cross Competentes*”, cuya restricción de cardinalidad es [1..1], es decir, una “*Cross Competentes*” puede ser agrupada en al menos una y sólo una “*Cross Unit*”.
- **Technical-Sectorial Unit:** Esta unidad corresponde a una reagrupación de las competencias técnico-sectoriales. Para dar una descripción detallada de éstas se debe analizar la descripción de su competencia asociada. Este nodo está relacionado al nodo “*Technical-Sectorial Competentes*”, cuya restricción de cardinalidad es [1..1], es decir, “*Technical-Sectorial Competentes*” puede ser agrupada en al menos una y sólo una “*Technical-Sectorial Unit*”.
- **Formative Path:** Este nodo hace referencia a la duración del curso, el título de todas las unidades que lo conforman, la duración en horas, el contexto de aprendizaje y los métodos de evaluación.
Indicar la resolución regional de aprobación del proyecto educativo constituye una información adicional y una garantía de transparencia. La duración del curso representa el atributo más inmediato que se puede considerar en una evaluación general de la coherencia entre los conocimientos y competencias que se esperan al momento de dar la certificación. Es decir, permite una evaluación indirecta de la calidad y confiabilidad del programa educativo. Este nodo se vincula al “*Final Certification*” a través de una restricción de cardinalidad [1..1], es decir, una certificación final detalla sólo un proceso formativo.
- **Formative Unit:** La unidad formativa corresponde a un módulo dentro del programa académico, por lo tanto para cada una de las unidades formativas se da su descripción en meses y horas. Más adelante será necesario especificar para cada uno de las unidades formativas el contexto de aprendizaje, es decir, si las

clases se realizaran en salas de clases, laboratorios, actividades individuales, proyectos, etc.

Además indicar los métodos de evaluación, entrevistas, pruebas escritas, exámenes prácticos, simulaciones, ejercicios y test, para así facilitar el reconocimiento de los créditos formativos que contiene cada curso dentro del programa. Todos estos conceptos mencionados se pueden clasificar bajo el concepto de unidad formativa interna.

Los descriptores de un curso pueden ser usados para destacar posibles entradas de créditos. En este caso es necesario especificar el título del curso que se reconoce como crédito, dejando fuera la duración en horas pero especificando el contexto de aprendizaje, las competencias que se han alcanzado y los métodos de verificación, en este caso se habla de unidad externa.

La unidad formativa es, por lo tanto, una generalización de la unidad formativa externa y de la unidad formativa interna. Existe una relación entre este nodo y el nodo “*Achieved Competentes*” a través de una restricción de cardinalidad [1..N], es decir, una unidad formativa es analizada en al menos una competencia.

- **Internal Formative Unit:** Como se mencionó antes este nodo corresponde a una especialización de la unidad formativa. Dentro de éste se almacenará la siguiente información, título de la unidad formativa, contexto de aprendizaje, métodos de evaluación, duración en horas y contenidos.
- **External Formative Unit:** Este nodo representa una especialización de un concepto más general que es la unidad formativa. Dentro de este nodo se almacenará la siguiente información: título de la unidad formativa, descripción del contexto de aprendizaje externo y métodos de evaluación.
- **Apprenticeship/Training:** Este nodo contiene información sobre el lugar en cual se realizará la práctica profesional, la duración, el sector o división donde se desempeñó el alumno, el proceso o función organizacional y las actividades que se realizaron. Esta información es importante porque ayuda a transparentar la calidad de la experiencia laboral del alumno. Este concepto está relacionado al “Formative Path” a través de una relación de cardinalidad [1..N].
- **Apprenticeship/Training Activity:** Este nodo contiene la descripción de las actividades que realizó el alumno en su periodo de práctica profesional. Este concepto está relacionado al concepto “*Apprenticeship/Training*” a través de una restricción de cardinalidad [1..N], es decir, al menos una o muchas actividades han sido realizadas dentro del “*Apprenticeship/Training*”.

- **Method for Final Evaluation:** En el certificado IFTS es necesario especificar la clase de métodos de evaluación que fueron realizados para medir los conocimientos y competencias alcanzados al final de la carrera.
Los exámenes finales en el sistema IFTS consisten en una interrogación individual y una prueba práctica. El certificado se entrega sólo si estas dos pruebas son aprobadas. El resultado de la evaluación se registra en el certificado sólo si el estudiante saca sobresaliente en sus exámenes y se registra como finalizado “con honores”.
Un certificado intermedio se entrega a aquellos estudiantes que no logran un resultado sobresaliente en sus exámenes. Este documento se estructura de acuerdo a un modelo definido por la comisión nacional para el diseño del IFTS e indica las competencias alcanzadas para facilitar el reconocimiento de los créditos formativos.
El método de evaluación final es, por lo tanto, una generalización del “*Individual Interview*” y del “*Simulation Test*”.
- **Individual Interview:** El examen oral se realiza según un expediente que contiene información sobre el proceso educativo del alumno, preparado por los profesores, contiene las evaluaciones de cada uno de los niveles por los que pasó el alumno así como los conocimientos culturales y competencias vocacionales que han sido adquiridas. Además de un documento individual preparado por el alumno que contiene una presentación, así como una reflexión personal del trabajo que ha realizado dentro de su periodo académico. El examen es preparado por la comisión, en conjunto con el comité técnico y científico. La calificación se da en notación centesimal, donde un 70% de nota corresponde al examen oral y al análisis del documento antes mencionado, y el 30% se atribuye a la prueba práctica. Un alumno aprueba si alcanza por lo menos 60 puntos entre los resultados de los dos exámenes.
Existe una relación que vincula este nodo con “*Method for Final Evaluation*” a través de una restricción de cardinalidad [1..1].
- **Simulation Test:** La prueba práctica permite verificar, la adquisición de conocimientos y competencias. El examen es preparado por la comisión, en conjunto con el comité técnico y científico. La calificación se da en notación centesimal, donde un 70% de nota corresponde al examen oral y al análisis del documento antes mencionado, y el 30% se atribuye a la prueba práctica. Un alumno aprueba si alcanza por lo menos 60 puntos entre los resultados de los dos exámenes. Existe una relación que vincula este nodo con el “*Method for Final Evaluation*” a través de una restricción de cardinalidad [1..1].

- **Formative Credit:** De acuerdo a las guías para el diseño de un curso IFTS, las competencias adquiridas dentro de ésta pueden ser reconocidas como un crédito formativo con respecto a otro sistema. Las unidades de crédito corresponden al peso en horas de clases que tiene una asignatura o en horas de trabajo práctico. Existen dos relaciones que vinculan este nodo con el “*Formative Path*” por un lado y con el “*Unit of Competence*”, en ambos casos la restricción de cardinalidad es [0..1].
- **Supplementary Annotation:** La información adicional se inserta, en caso de necesidad, para agregar los detalles cualitativos de la experiencia y de la modalidad en la cual se ha realizado el programa educativo. Este nodo está vinculado al “*Final Certificate*” a través de una restricción de cardinalidad [0..N], es decir una certificación final puede incluir cero o muchas “*Supplementary Annotation*”.

La Figura 4.10 muestra la red semántica del certificado IFTS.

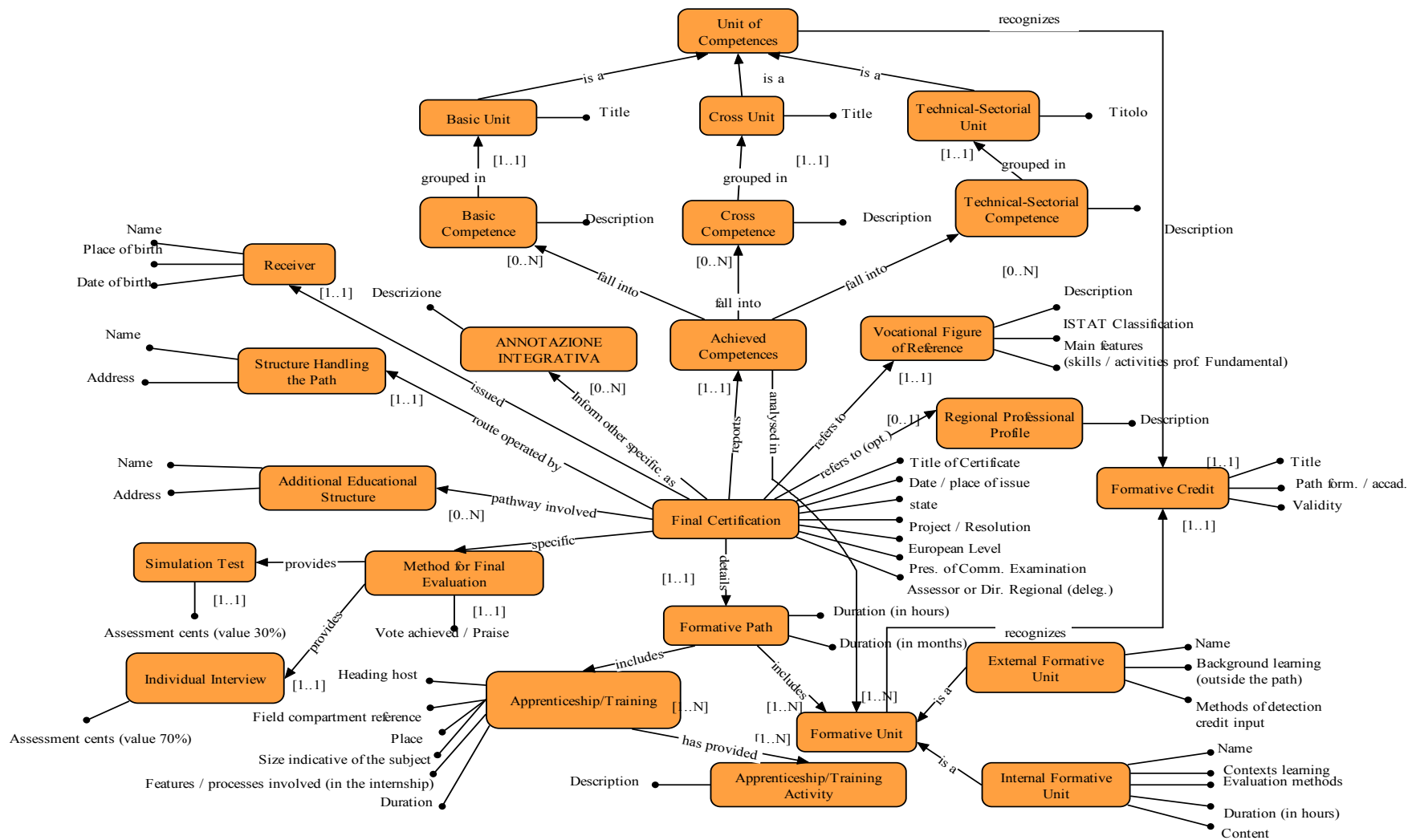


Figura 4.10. Red Semántica del Certificado IFTS.

4.5 Irlanda

4.5.1 Introducción

En los siguientes párrafos se describirá la certificación en Irlanda, este país se seleccionó como caso de estudio sólo para poder ampliar los conocimientos en cuanto a la certificación en cada país de Europa y ver las diferencias existentes entre un país y otro. Con ello se logró obtener una red semántica consistente del certificado irlandés FETAC que se describirá en este punto.

No se utilizará un programa de estudio irlandés para comprobar la utilidad de Information Retrieval en la comparación semántica de certificados, debido que no se encontraron disponibles casos de estudios para poder efectuar las pruebas.

4.5.2 Certificación en Irlanda

La certificación en Irlanda depende de la Autoridad Nacional de las Cualificaciones de Irlanda “*National Qualification Authority of Ireland NQAI*” [7], fue establecida en el artículo del 26 de febrero del 2001 bajo cualificaciones en educación y entrenamiento. Los objetivos principales de esta institución son asegurar que todo el entrenamiento vocacional y educativo realizado dentro del estado esté reconocido y estandarizado tanto en calidad como en educación, además de establecer y mantener un framework nacional de cualificaciones junto con promover y facilitar el acceso, la transferencia y el aumento de la educación en Irlanda.

Esta autoridad tiene dos organizaciones relacionadas: el “*Further Education and Training Council Awards (FETAC)*” y el “*Higher Education and Training Council (HETAC)*” que son responsable de acreditar todos los cursos educativos y vocacionales así como aquellos cursos certificados por cualquier universidad o por el instituto de Tecnología de Dublín.

FETAC [29] es responsable de todos los cursos vocacionales y de educación superior hasta el sexto nivel de certificación, es decir, concede certificados a la educación técnica no-universidad. Para cada nivel de certificación el estudiante debe satisfacer ciertos estándares detallados a continuación:

Nivel 1, Los resultados del aprendizaje se relacionan con la ejecución de tareas básicas en un ambiente controlado bajo la supervisión y exhibición de habilidades básicas y repetitivas, así como la secuencia de tareas de aprendizaje.

Nivel 2, Los resultados claves de este nivel son el conocimiento y la capacidad de cálculos básicos y la introducción a un aprendizaje sistemático. Los resultados del aprendizaje se relacionan con la capacidad de aprender nuevas habilidades y conocimientos en un ambiente supervisado y de realizar un trabajo rutinario bajo dirección. Los resultados del aprendizaje a este nivel son típicamente de desarrollo más que relacionados a una ocupación específica.

Nivel 3, Los resultados del aprendizaje de este nivel se relacionan con un bajo volumen de capacidad práctica y de conocimiento teóricos, ejecución de trabajos relativamente simples y que pueden ser aprendidos bastante rápido. Éstos también pueden conceder una posibilidad de empleo mínimo para ocupaciones de habilidades simples que incluyen el conocimiento y capacidades de cálculos funcionales básicos.

Nivel 4, La independencia es el sello de este nivel. Los resultados del aprendizaje corresponden a un sentido cada vez mayor de responsabilidad para participar en la vida pública y formar su propia vida. Los resultados a este nivel están asociados a la entrada en muchos sectores ocupacionales.

Nivel 5, Los resultados del aprendizaje a este nivel incluyen una amplia gama de las habilidades que requieren una cierta comprensión teórica. Los resultados pueden relacionarse con el vínculo a una actividad específica, con la capacidad de utilizar instrumentos y las técnicas referentes a una ocupación. Se asocian al trabajo que es realizado en forma independiente, conforme a una dirección general.

Nivel 6, Los resultados del aprendizaje a este nivel incluyen una gama comprensiva de las habilidades que pueden ser vocacionales, específicas y/o de una naturaleza de supervisión general, y requieren la comprensión teórica detallada. Los resultados se relacionan con el trabajo de una manera generalmente autónoma de asumir diseño y/o gerencia y/o responsabilidades administrativas.

FETAC es responsable de desarrollar un nuevo sistema de cualificaciones para la enseñanza superior y los sectores de entrenamiento. Este sistema es parte del desarrollo completo del nuevo framework de 10 niveles de cualificaciones que incluyen todas las entregas disponibles desde el nivel más básico (1) hasta el nivel más avanzado (10) de educación. FETAC tiene la responsabilidad específica de proporcionar una gama de entregas desde los niveles 1 al 6 dentro del nuevo framework.

Dentro del área de estudios que es el sector técnico no universitario los certificados irlandeses correspondientes a este grupo corresponden al nivel 6 dentro de su framework educacional. Por lo tanto, el modelo del certificado que se muestra a continuación

corresponde al caso de estudio y será formalizado a través de una red semántica. La figura 4.11 muestra el certificado entregado por FETAC.

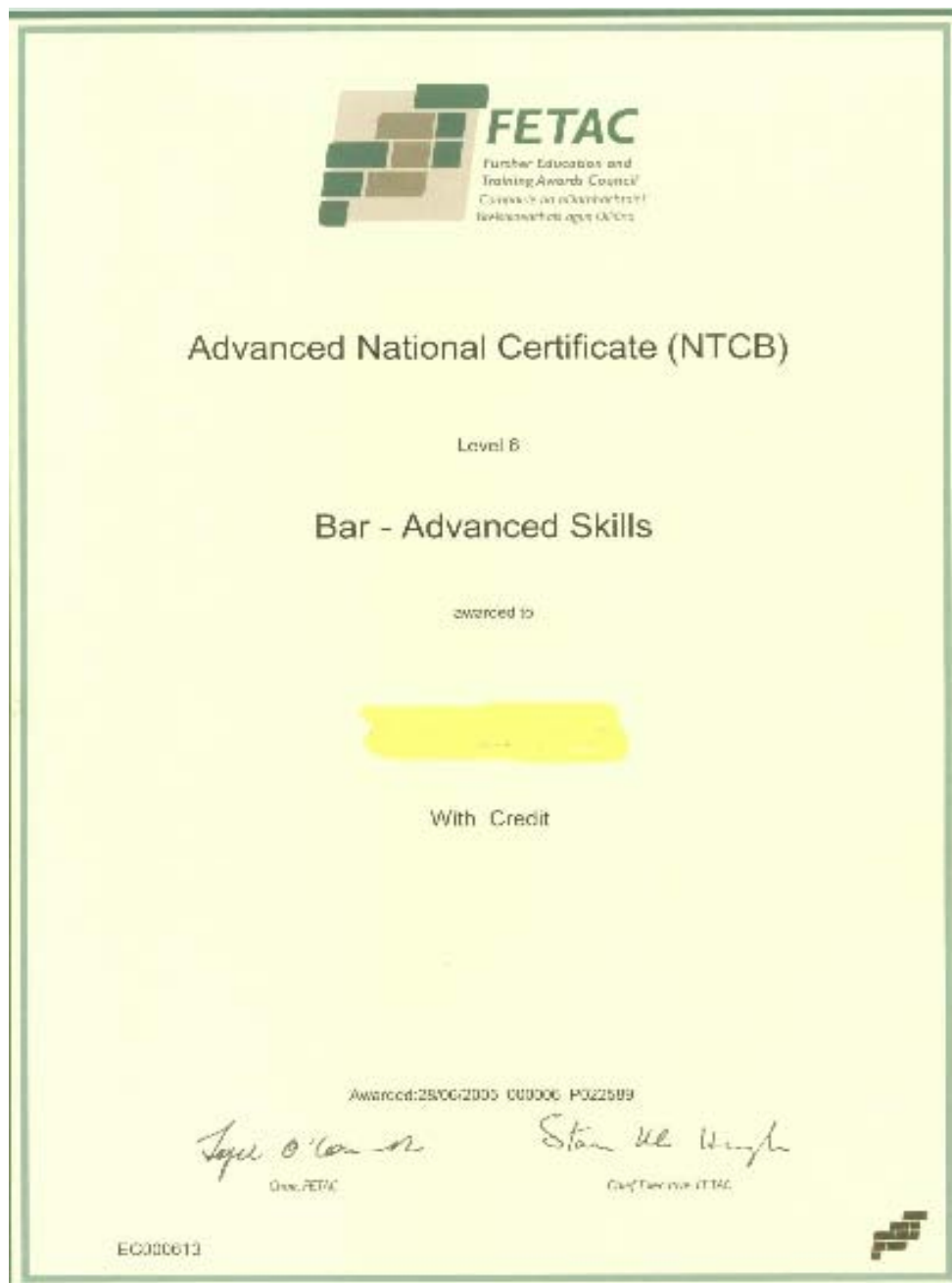


Figura 4.11. Formato del certificado FETAC.

4.5.3 Descripción de la Red Semántica del Certificado FETAC

- **FETAC Advanced Certificate:** La certificación FETAC incluye información detallada sobre el Título del certificado, el nivel en el sistema educacional irlandés, el día de entrega y la autoridad responsable que firma el documento. Por lo tanto, el concepto de “*FETAC Advanced Certificate*” ha sido modelado como un nodo complejo.
- **Receiver:** Es necesario tener un cuidado especial en que este ítem entregue sólo información necesaria. Para evitar que cualquier característica personal incluida en el documento se pueda utilizar o interpretar como un elemento de discriminación con respecto a la posibilidad de acceder a un trabajo. Consecuentemente esta información asociada al destinatario se limitará a los elementos necesarios para su identificación.
Este concepto se ha modelado con un nodo complejo que agrupa otro nodo básico u hoja y está estrictamente relacionado al nodo “*FETAC Advanced Certificate*”. La relación se establecerá a través de una flecha que lo conecte con el concepto antes mencionado dentro de la red semántica. Esta relación tendrá una restricción de cardinalidad [1,1]. Esto significa que un “*FETAC Advanced Certificate*” es entregado por lo menos un a “*Receiver*”.
- **NQAI:** Es la institución a cargo de la certificación en Irlanda. Este concepto se ha modelado como nodo complejo que agrupa varios nodos básicos u hojas. Existe una relación que vincula los conceptos de “*NQAI*” y de “*NFQ*” que se muestra dentro del modelo como una flecha que junta ambos conceptos, el nombre de la relación es la trayectoria “*se maneja por*”.
- **NFQ:** Hay un número de cambios importantes que ocurrieron en el sistema de cualificación en Irlanda. El más significativo es la introducción del marco nacional de las cualificaciones (*Nacional Framework Qualification NFQ*). El NFQ es un sistema de diez niveles.
Se diseñó para incorporar las entregas hechas a todas las clases de aprendizaje donde quiera que este se haya obtenido. Pone las necesidades del estudiante primero y apoya el objetivo nacional de movilidad hacia una “sociedad de aprendizaje de por vida”.
El framework irlandés se ha estructurado de la siguiente manera dentro de los 10 niveles reconocidos: nivel 10 doctor, nivel 9 grado de magíster y diploma de postgrado, nivel 8 grado de bachelor con honores y diploma superior, nivel 7 grado de bachelor, nivel 6 certificado avanzado y certificado superior, nivel 5

certificado nivel 5, nivel 4 certificado nivel 4, nivel 3 certificado nivel 3, nivel 2 certificado nivel 2, nivel 1 certificado nivel 1.

Este concepto se ha modelado como nodo complejo que agrupa varios nodos básicos u hoja. Existe una relación que vincula el “*NFQ*” al “*NFQ Levels*” que se muestra en el modelo como una flecha que une ambos conceptos; su nombre es la trayectoria “*se divide en*” de NFQ hacia el NFQ levels.

- **NFQ Levels:** La autoridad ha determinado que el framework consistirá en 10 niveles. Los niveles del framework precisan una gama de estándares de conocimiento, habilidades y competencias. Ocho sub-estándares de habilidades, competencias y conocimientos han sido identificados por la autoridad. Éstos se precisan detalladamente en las políticas y criterios de comparación.
- **Certificate Level:** Éste es un nodo complejo que contiene la información con respecto al nivel de un FETAC Advanced Certificate dado por la institución del mismo nombre, para validar los conocimientos y las habilidades del poseedor del certificado. En este sentido, el nivel corresponde al sexto dentro del NFQ. En términos prácticos este concepto corresponde a una especialización del concepto del nivel de NFQ, razón por la que se ha modelado como nodo complejo. Existe una relación que vincula los conceptos “*Certificate Level*” y “*NFQ Level*” que se muestra en el modelo como una flecha que une ambos conceptos. Su nombre es la trayectoria “*es un*”. Además, hay otra relación que conecta el “*Certificate Level*” con el concepto “*FETAC*”. En este caso el nombre de la trayectoria se denomina “*es responsable de*”.
- **Next Level:** Éste es un nodo complejo que contiene la información con respecto al siguiente nivel de educación al que puede tener acceso el estudiante cuando su trayectoria formativa ha acabado. Existe una relación que une el “*Next Level*” y el concepto “*NFQ Level*” que se muestra en el modelo como una flecha que junta ambos conceptos. El nombre de esta relación corresponde a “*es un*”. Además, hay otra relación que conecta el “*Next Level*” con el concepto de “*HETAC*”. En este caso el nombre corresponde a “*es responsable de*”.
- **FETAC:** Es la institución nacional para la enseñanza y el entrenamiento en Irlanda. FETAC tiene la responsabilidad de hacer y promover entregas hechas previamente por BIM, Fáilte Irlanda, FÁS, NCVA y Teagasc. La misión de FETAC es asegurar la calidad de las entregas de acuerdo con los estándares nacionales dentro del framework nacional. FETAC da a las personas la oportunidad de reconocer el aprendizaje obtenido en centros de educación o entrenamiento y en el lugar de trabajo. Este concepto

se ha modelado como un nodo complejo que agrupa varios nodos básicos u hojas. Existe una relación que une el “*Certificate Level*” y el concepto “*FETAC*” y esta relación se muestra en el modelo como una flecha que junta ambos conceptos; su nombre es la trayectoria “*es responsable de*”.

- **HETAC:** El “*Higher Education and Training Awards Council*”, fue establecido el 11 de junio de 2001, bajo el artículo 1999 de las cualificaciones (educación y entrenamiento). Es la institución encargada de reconocer los estudios a nivel universitario. Este concepto tendrá información adicional tal como el nombre, el E-mail y el número de teléfono. Por esta razón, el concepto se ha modelado como nodo complejo que agrupa varios nodos básicos u hojas. Existe una relación que une el “*Next Level*” y el concepto “*HETAC*”, y se muestra en el modelo como una flecha que conecta ambos conceptos juntos; su nombre es la trayectoria etiquetada como “*es responsable de*”.
- **Programme:** El programa contiene la información sobre el curso que va a ser tomado por el estudiante, por lo tanto, se modela como nodo complejo dentro de la red semántica, porque agrupa varios nodos básicos u hojas que describan características diversas tales como: el título del curso, el perfil de éste, la duración, estructura, objetivos, criterios profesionales de admisión, prácticas profesionales así como la carga y la certificación. Existen dos relaciones que unen el concepto “*Programme*”, a los conceptos “*Profile Statement*” y “*FETAC Advanced Certificate*”. Estas relaciones se muestran en la red semántica como una flecha que une estos conceptos. La primera conecta el concepto “*Programme*” con el “*Profile Statement*” con la etiqueta “*contiene*” y la segunda vincula el concepto “*FETAC Advanced Certificate*” al concepto “*Programme*” por medio de una flecha etiquetada como “*se relaciona con*”.
- **Profile Statement:** Son los conocimientos y habilidades que deben desarrollar los estudiantes que se gradúan de los programas académicos y que les permiten incorporarse al mercado laboral satisfaciendo los requisitos profesionales que éste exige. Este se ha modelado en la red semántica como un nodo complejo.

La figura 4.12 muestra la red semántica del certificado FETAC.

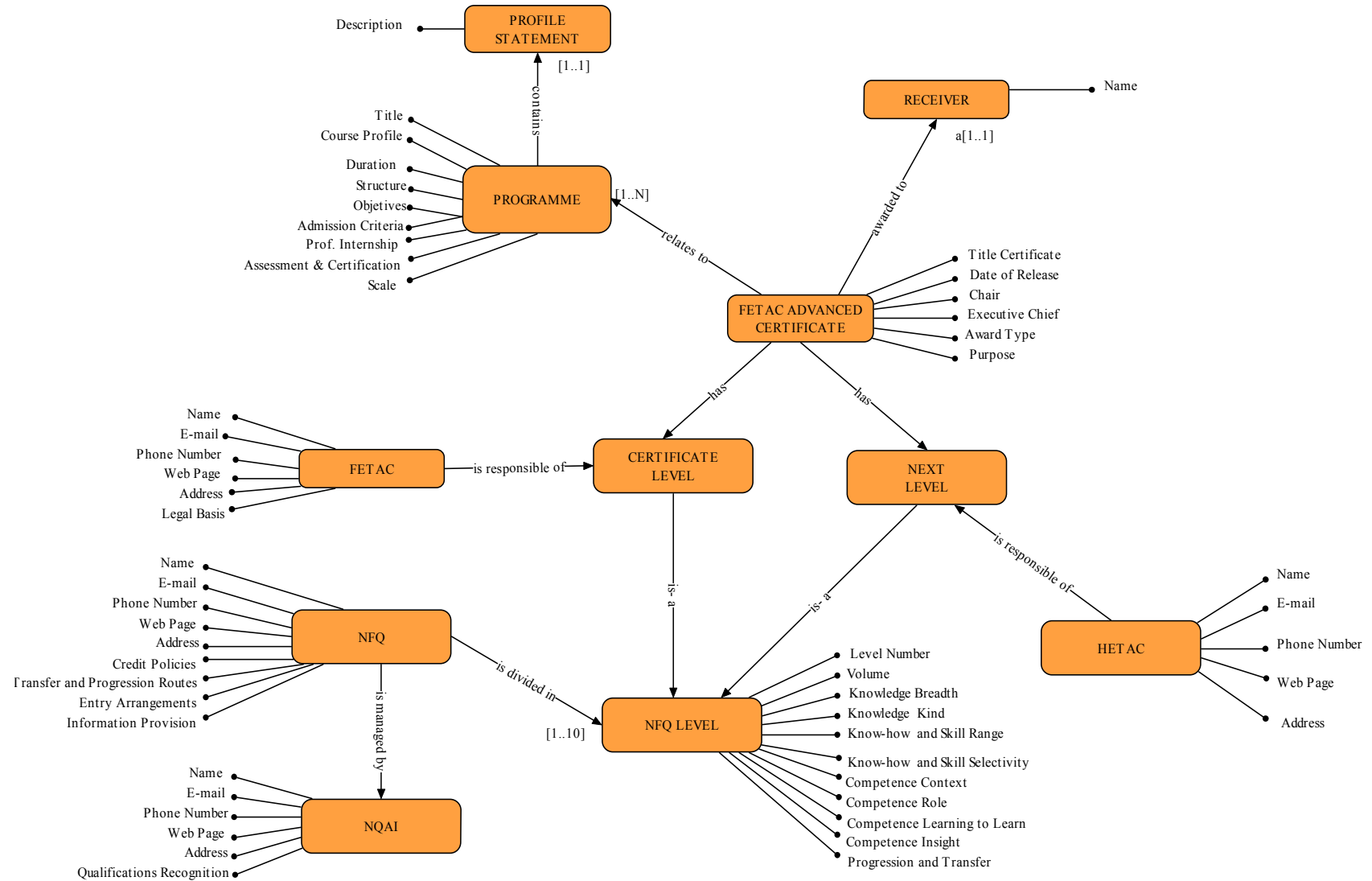


Figura 4.12. Representación formal de la red semántica del certificado FETAC.

4.6 Holanda

4.6.1 Introducción

A continuación se describirá el proceso de certificación en Holanda, el último país seleccionado como caso de estudio tanto para poder ampliar la visión en cuanto a lo que es la certificación en Europa además de utilizar un programa académico perteneciente a este país para comprobar la utilidad de Information Retrieval en la comparación semántica de competencias con el fin poder determinar si dos programas académicos de distintos países pueden ser homologados.

Se seleccionó Holanda como caso de estudio debido a la cantidad de información disponible con lo que se pudo obtener un modelo semántico consistente, por lo tanto, dentro de este punto también se describirá la red semántica generada a partir de certificado holandés.

4.6.2 Certificación en Holanda

La educación vocacional es un proveedor clave de profesionales calificados para los negocios y la sociedad. La mayoría de los jóvenes holandeses siguen un curso de educación vocacional. Las bases de esta educación se confeccionan en función de los negocios y la industria [10].

El framework educacional holandés está compuesto de cuatro completos niveles educacionales:

- **Educación Primaria:** Los colegios primarios en Holanda educan niños desde los cuatro a los doce años. Los ocho primeros años el currículum escolar se basa en el desarrollo emocional, intelectual y creativo del niño, junto con la entrega de habilidades sociales, culturales y físicas. Cada colegio primario realiza su propio plan de trabajo escolar basado en las pautas dadas por el gobierno.
- **Educación Secundaria:** Los niños a partir de los doce años pueden ingresar a colegios secundarios, dentro de los cuales se encuentran tres tipos:
 - ✓ Educación Secundaria Pre-vocacional (VMBO).
 - ✓ Educación Secundaria General Senior (HAVO);
 - ✓ Educación Pre-universitaria (VWO).

La mayoría de los colegios secundarios proporcionan más de un tipo de educación. VMBO toma cuatro años y permite que el alumno acceda a la educación vocacional secundaria MBO.

HAVO es un programa de cinco años y permite continuar a la educación profesional superior HBO.

VWO consiste de un programa de seis años y permite al joven acceder a la educación universitaria.

Los alumnos de colegios secundarios los primeros dos a tres años de estudios tienen un currículo base que consiste de 15 cursos.

- **Educación Vocacional:** En 1996 se establecieron centros regionales de preparación que ofrecen una amplia gama de cursos para la educación vocacional y adulta. Todos los cursos combinan el trabajo con el estudio.

Hay dos tipos de programas: preparación vocacional con al menos un 20% de trabajo práctico y no más de un 60% y el otro tipo considera un trabajo práctico de por lo menos un 60%. Estos cursos se imparten a través de cuatro niveles.

- **Educación Superior:** Se enfoca a los profesionales universitarios. En la estructura del framework educacional danés en el 2002 se introdujo el grado de “Bachelor-Master”. Para obtener el grado “Bachelor” se debe completar un curso de tres años universitarios. Los graduados con este grado pueden optar al grado de Master tomando un curso de uno a dos años de duración, dependiendo del área.

Dentro del framework los certificados entregados a alumnos de nivel técnico no universitario corresponden al cuarto nivel. Las instituciones que entregan educación de este tipo se rigen por la estructura de cualificación “LOB HTV” es un sistema de cualificación vocacional reconocida y regida por el ministerio de educación, cultura y ciencia. Este sistema educa jóvenes dentro del área de hospitalidad (hoteles, restorán, cafés), cocina, encargado de bodega, sector turístico y panadería.

LOB HTV es una colaboración entre compañías y colegios que entregan educación técnica en los sectores antes mencionados que trabajan en forma coordinada para que los programas de estudio satisfagan las necesidades del mercado laboral. El sistema LOB HTV estimula y mejora la calidad de los establecimientos educacionales.

4.6.3 Descripción del Certificado Holandés

Los centros de experiencia en educación vocacional, entrenamiento y mercado laboral como LOB HTV representa el vínculo entre la educación y los sectores organizados en negocios e industrias. La mayoría de los centro de experiencia se establecieron para organizar los sectores de negocios e industriales, y se organizan de acuerdo a la categoría

del sector o profesión. La tarea de estos centros es traducir los nuevos y rápidos cambios de los requerimientos del mercado laboral en la práctica profesional dentro del formato educacional. La figura 4.13 muestra el certificado entregado por el sistema danés de cualificación.



Figura 4.13. Formato del Certificate of Qualification.

4.6.4 Descripción de de la Red Semántica del Certificado de Cualificación

- **Certificate of Qualification:** El certificado de cualificación incluye información detallada con respecto al título del certificado y la autoridad competente responsable de entregarlo. Por lo tanto, el concepto “*Certificate of Qualification*” se ha modelado como nodo complejo.

- **Receiver:** Es necesario tener especial cuidado que este ídem entregue sólo información necesaria. Esto es para evitar que cualquier característica personal incluida en el documento se pueda utilizar o interpretar como elemento de discriminación con respecto a la posibilidad de entrar a un trabajo. Por lo tanto, esta información asociada al receptor se limitará a los elementos necesarios para su identificación.

Este concepto se ha modelado con un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con “*Certificate of Qualification*” a través de una flecha que conecta ambos conceptos en la red semántica. Tiene una restricción de cardinalidad [1, 1]. Esto significa que un “*Certificate of Qualification*” es entregado a uno y sólo un receptor.

- **List of Grades:** Corresponde a la lista de grados que se distribuye después de la examinación, en la cual entre otras cosas detalla, la comisión docente que participó de la examinación y la comisión del centro de experiencia a cargo de la evaluación del estudiante, la lista describe los temas académicos que fueron evaluados y el resultado de esta evaluación. Además da la información sobre el proceso de estudios, el nombre del estudiante, nivel de estudios dentro del sistema educativo holandés, la cualificación, duración del curso, la escala de calificación, la fecha y el lugar de la entrega del certificado.

Este concepto se ha modelado con un nodo complejo que agrupa otros nodos básicos u hojas. Se relaciona con el “*Certificate of Qualification*”, a través de una relación que conecta ambos conceptos en la red semántica. Esta relación es caracterizada por una restricción de cardinalidad [1,1]. Esto significa que un “*Certificate of Qualification*” es completada por una y sólo una “*List of Grades*”.

- **Explanation Vocational Education:** Como se mencionó anteriormente la educación vocacional es un proveedor clave de las cualificaciones profesionales para el negocio y la sociedad.

La mayoría de los jóvenes holandeses siguen cursos de educación vocacional. Este concepto se relaciona con el “*Certificate of Qualification*”, a través de una

relación dentro de la red semántica. Esta relación es caracterizada por una restricción de cardinalidad [1,1], lo que significa que un “*Certificate of Qualification*” se relaciona por lo menos con una y sólo una “*Explanation Vocational Education*”.

- **Professional Practice:** Es importante destacar que la mayor parte de la educación vocacional holandesa sucede en la práctica profesional. Por lo tanto, los centros de experiencia supervisan que haya suficientes establecimientos de trabajo que participen en los programas educativos.

La información dentro de este concepto se refiere al período obligatorio de práctica profesional que debe satisfacer un estudiante para terminar su formación académica.

Para el nivel educacional en estudio que es el técnico no universitario, un estudiante debe realizar dos períodos de práctica profesional de veinte semanas cada una.

Este concepto se ha modelado con un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el “*Explanation Vocational Education*” a través de una relación dentro de la red semántica. Esta relación es caracterizada por una restricción de cardinalidad [1, 2]. Esto significa que un “*Explanation Vocational Education*” debe tener por lo menos una y hasta dos prácticas profesionales.

- **Centre of Expertise:** Los centros de experiencia comparten conjuntamente la responsabilidad por la educación vocacional de calidad. Forman parte de una red internacional de organizaciones.

Cada vez más, estos centros de experiencia colaboran con sus socios extranjeros, por ejemplo en el reclutamiento de los establecimientos de aprendizaje en el exterior, y la comparación internacional de las cualificaciones. Las tres tareas base de estas instituciones son: la innovación y el mantenimiento de la estructura de cualificación, la supervisión y promoción de la calidad de aprendizaje en la práctica profesional y la responsabilidad común en la calidad de las exámenes.

Este concepto se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el concepto “*List of Grades*”, a través de una relación dentro de la red semántica con una restricción de cardinalidad [1, N]. Esto significa que un “*List of Grades*” está relacionado por lo menos con uno y a lo más N “*Centre of Expertise*”.

- **Colo:** Es la agencia a cargo de coordinar los centros de la experiencia, tiene el deber de desarrollar y mantener la estructura de cualificación holandesa,

garantizando una alta calidad y una amplia gama de colocaciones para la práctica.

Este concepto se ha modelado con un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el concepto “*Centre of Expertise*” a través de una relación dentro de la red semántica con una restricción de cardinalidad [1, N]. Esto significa que la agencia “*Colo*” está relacionada por lo menos con una y a lo más N “*Centre of Expertise*”.

- **School:** Los colegios corresponden a la institución académica a cargo de la formación de los ciudadanos holandeses.

También junto con representantes de la industria toman las exámenes que aseguran la calidad profesional de los estudiantes que se incorporan al mercado laboral.

Este concepto se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el concepto “*List of Grades*” y con el concepto “*Explanation Vocational Education*”, a través de una relación dentro de la red semántica. Estas relaciones son caracterizadas por restricciones de cardinalidad [1, 1]. Esto significa que un “*School*” está relacionado por lo menos a una y sólo una “*List of grades*”. Por otro lado, un “*School*” contiene por lo menos una y sólo una “*Explanation Vocational Education*”.

- **Party Qualification:** Las cualificaciones son el conjunto de competencias que serán evaluadas por las comisiones escolásticas y las comisiones de los centros de experiencias en la lista de grados y que determinan la calidad profesional del estudiante.

Cada una de estas competencias se evalúa en una escala de calificación de 1-10 donde la nota mínima de aprobación para cada competencia debe ser 5,5.

Este concepto se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el “*List of Grades*”, a través de una relación que lo conecta con este concepto en la red semántica. Esta relación es caracterizada por una restricción de cardinalidad [1, N]. Esto significa que un “*List of Grades*” contiene por lo menos una y a lo más N “*Party Qualification*”.

- **Qualification Profile:** El perfil de cualificaciones es la base de la estructura de calificación que describe el sistema educativo holandés.

El perfil de cualificaciones describe las competencias profesionales de un trabajador que recién comienza en el mercado laboral, junto con las actividades del contexto profesional en el cual se desempeñará. Los perfiles de cualificaciones se elaboran de las tareas claves existentes, problemas existentes

y competencias, todas presentes en las necesidades del mercado laboral holandés.

Este concepto se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el concepto “*Certificate of Qualification*” a través de una relación dentro de la red semántica con una restricción de cardinalidad [1, 1]. Esto significa que un “*Certificate of Qualification*” tiene por lo menos uno y sólo un “*Qualification Profile*”.

- **Core Duty:** Este concepto conceptualiza las competencias, es decir, define el alcance del trabajo que se deberá realizar y describe como debe ser hecho. Por lo tanto, se ha modelado con un nodo complejo que agrupa otro nodo básico u hoja.

Este concepto se relaciona con el concepto de “*Profile Qualification*” a través de una relación dentro de la red semántica con una restricción de cardinalidad [1, N]. Esto significa que un “*Profile Qualification*” tiene por lo menos uno y a lo más N “*Core Duty*”.

- **Core Assignment:** Este concepto describe diversas situaciones con las cuales se encontrará el estudiante en su vida profesional, dando pautas para que enfrente estas situaciones de la mejor manera. Por lo tanto, este concepto se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja.

Se relaciona con el concepto “*Qualification Profile*” a través de una relación dentro de la red semántica con una restricción de cardinalidad [1, N]. Esto significa que un “*Qualification Profile*” contiene por lo menos uno y a lo más N “*Core Assignment*”.

- **Competences:** Este concepto contiene la información detallada de las competencias que debe desarrollar el estudiante a través del programa educativo que está siguiendo. Por lo tanto, se ha modelado como un nodo complejo que agrupa otro nodo básico u hoja. Se relaciona con el concepto “*Qualification Profile*” a través de una relación con una restricción de cardinalidad [1, N]. Esto significa que un “*Qualification Profile*” contiene por lo menos una y a lo más N “*Competentes*”.

La figura 4.14 muestra la red semántica del Certificate of Qualification.

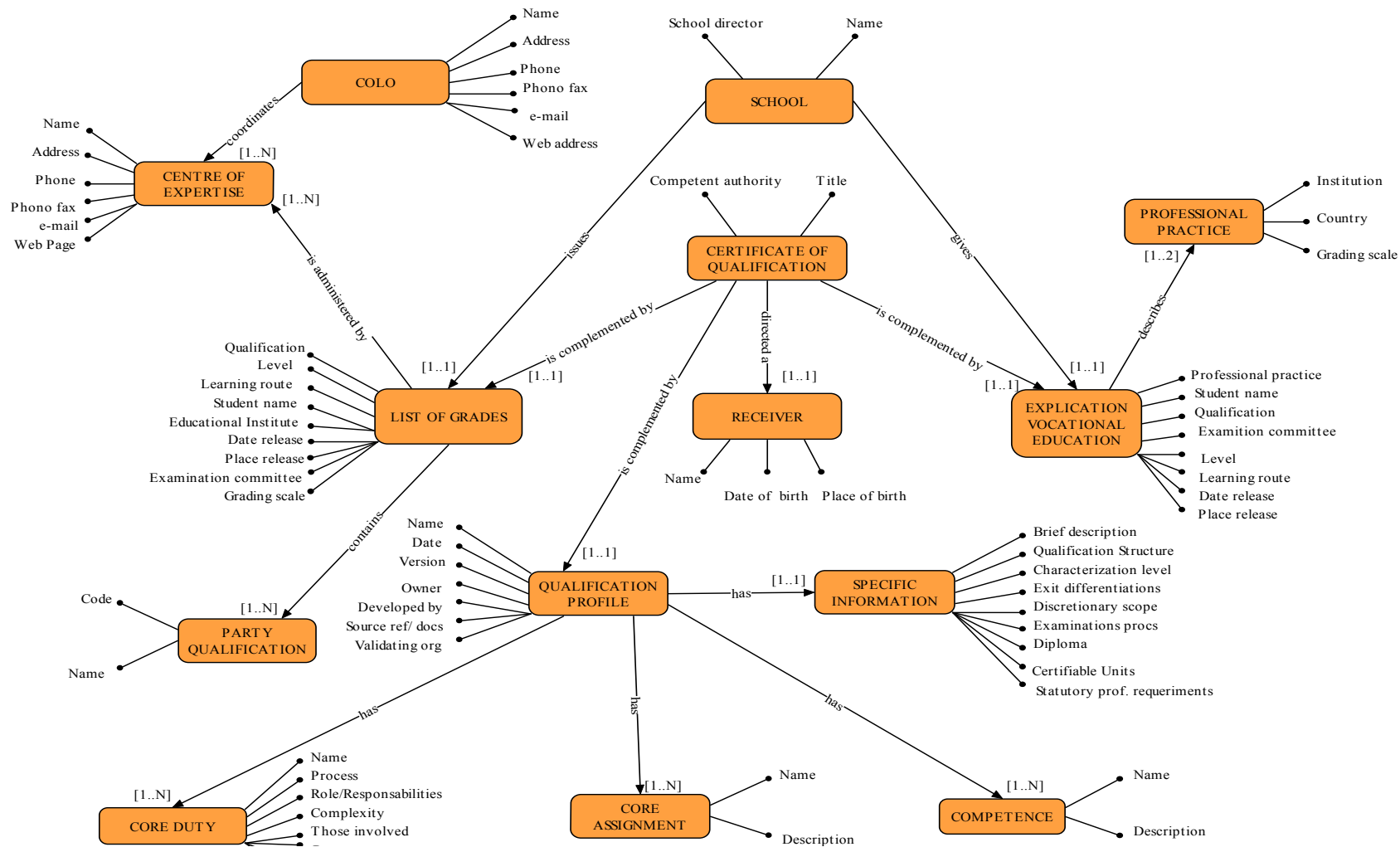


Figura 4.14. Representación formal de la Red Semántica del Certificate of Qualification.

4.7 Reglas de Mapping

Para poder realizar la integración de los distintos certificados, fue necesario establecer qué campos de cada red semántica de los países analizados en los puntos anteriores entregan la información necesaria para llenar cada una de las instancias del meta modelo C.S y con ello lograr tener un modelo formal de los diferentes certificados de formación técnica.

Obtener la información para llenar cada uno de los puntos del C.S con la información proporcionada por los modelos semánticos descritos en los puntos anteriores es relativamente sencillo a excepción de la sección de competencias donde no hay uniformidad para describir esta información por lo tanto se hace necesario utilizar varios nodos de la red semántica de cada país para poder completarla.

Por lo descrito en el párrafo anterior es que se hace necesario utilizar técnicas de Information Retrieval para poder realizar una comparación de competencias que determine si un plan de estudios de una determinada carrera técnica de un país puede ser homologado para la misma carrera o similar en otro país.

La tabla 4.1 muestra el mapping entre cada uno de los nodos de la red semántica del C.S con los de las redes de Italia y Holanda, establecer las reglas de mapping fue posible luego de analizar detalladamente la información que es almacenada en cada nodo de las distintas redes semánticas de los países descritos en los puntos anteriores. En muchos casos fue necesario utilizar más de un nodo para poder completar un punto del C.S. porque así se puede obtener una idea más clara y completa de la información que se requiere que describa ese campo.

Tabla 4.1. Mapping entre el Certificate Supplement y los países Italia y Holanda.

ITALIA	C.S	HOLANDA
<ul style="list-style-type: none"> - Final certification - Vocational figure of reference - Regional professional profile 	1. Certificate Supplement (Certificate original and translated title, nation, national reference point URL)	<ul style="list-style-type: none"> - Certificate of Qualification. <li style="padding-left: 20px;">- Colo. - Qualification Profile. - Specific Information. <li style="padding-left: 20px;">- List of Grades. - Vocational Education Explanation.
<ul style="list-style-type: none"> - Vocational figure of reference - Regional professional profile <li style="padding-left: 20px;">- Achieved competences <li style="padding-left: 20px;">- Basic competence 	2. Profile of skills and competences	<ul style="list-style-type: none"> - Qualification Profile. - Competences. <li style="padding-left: 20px;">- Core Duty. <li style="padding-left: 20px;">- Core Assignment.

<ul style="list-style-type: none"> - Cross competence - Technical-sectorial competence - Units of competence <ul style="list-style-type: none"> - Basic Unit - Cross Unit - Technical-sectorial Unit 		<ul style="list-style-type: none"> - List of Grades. - Qualification Party.
<ul style="list-style-type: none"> - Vocational figure of reference - Regional professional profile <ul style="list-style-type: none"> - Units of competence - Technical-sectorial Unit - Formative credit 	<p>3. Range of occupations accessible to the holder of the certificate</p>	<ul style="list-style-type: none"> - Certificate of Qualification. - Qualification Profile. <ul style="list-style-type: none"> - Core Duty. - Core Assignment. - Competences. - Professional Practice.
<ul style="list-style-type: none"> - Final certification - Vocational figure of reference - Regional professional profile - Structure handling the path <ul style="list-style-type: none"> - Additional educational - Method for final evaluation 	<p>4. Official basis of the certificate (Awarding body, national /regional authority, level, grading scale, next level of education, international agreements, legal basis)</p>	<ul style="list-style-type: none"> - Colo. - Centre of Expertise. - School. - List of Grades. - Vocational Education Explanation. - Qualification Profile. - Specific information. - Certificate of Qualification
<ul style="list-style-type: none"> - Formative path - Formative Unit <ul style="list-style-type: none"> - Internal formative Unit - External formative Unit - Apprenticeship/training - Method for final evaluation 	<p>5. Officially recognized ways of acquiring the certificate (type of received education and training, total duration)</p>	<ul style="list-style-type: none"> - List of Grades. - Vocational Education Explanation. - Professional Practice
<ul style="list-style-type: none"> - Final certification - Vocational figure of reference - Regional professional profile <ul style="list-style-type: none"> - Formative credit 	<p>6. Entry requirements (Education level, mode of selection, etc)</p>	<ul style="list-style-type: none"> - Qualification Profile. - Specific Information.
<ul style="list-style-type: none"> - Apprenticeship/training - Apprenticeship/Training - Method for final evaluation <ul style="list-style-type: none"> - Individual interview - Simulation test - Supplementary annotation 	<p>7. Additional information (type of practical training, programme or structure of the vocational training system, assessment procedure, etc)</p>	<ul style="list-style-type: none"> - Qualification Profile. - Specific Information - Vocational Education Explanation. - List of Grades.
<ul style="list-style-type: none"> - Supplementary annotation 	<p>8. More information URL (with a further description of the national qualifications system)</p>	<ul style="list-style-type: none"> - Colo. -School. - Centre of Expertise.

Capítulo 5

Desarrollo de la Aplicación Ilustrativa

5.1 Introducción

Dentro de este capítulo se detallará el proceso de desarrollo de la aplicación ilustrativa que se utilizará para comprobar la efectividad y validar así la técnica de Information Retrieval seleccionada que corresponde al algoritmo del Clasificador TFIDF explicado en el capítulo 2.

Como primer punto se describirán los requerimientos tanto funcionales como no funcionales necesarios para poder desarrollar la aplicación. Además de diagramas de casos de uso y de secuencia que serán implementados para especificar algunas de las funcionalidades de la aplicación ilustrativa que se encargará de aplicar la técnica de Information Retrieval seleccionada. Además de describir como se utilizó ésta para lograr obtener el grado de similitud entre competencias que determinan si un programa académico de un país puede ser homologado en otro.

También se expondrán los modelos entidad relación diseñados a partir de las redes semánticas generadas y mostradas en el capítulo anterior. Además se describirán los casos de estudio de los países seleccionados correspondientes a Italia y Holanda, para aplicar la técnica de Information Retrieval. Con estos modelos se creará la base de datos en la cual se almacenará la información necesaria para cumplir con los requerimientos de la aplicación.

Se consideró utilizar sólo Italia y Holanda para esta parte del proyecto debido que se encontraban a disposición los planes académicos de las carreras seleccionadas y la cantidad de información que contenían era suficiente para poder comprobar si el Clasificador TFIDF perteneciente a Information Retrieval es útil para poder hacer una comparación semántica de competencias.

Finalmente se presentarán las interfaces de la aplicación implementada para así tener una mejor comprensión de la navegabilidad de ésta. Para mostrar por último los resultados obtenidos.

5.2 Requerimientos de la Aplicación ilustrativa

Los *Requerimientos Funcionales* de la aplicación ilustrativa servirán de apoyo para mostrar los resultados obtenidos al aplicar el algoritmo de Information Retrieval seleccionado para los casos de estudio escogidos. A continuación se describen éstos:

- La aplicación debe ser capaz de entregar el grado de similitud entre competencias y habilidades para una pareja de carreras de dos países diferentes haciendo uso de las técnicas de Information Retrieval, obteniéndose un ranking con los grados de similitud entre las competencias.
- La aplicación debe ser capaz de generar un documento con el formato del Certificate Supplement para mostrar la información de cada país de manera uniforme y estándar al experto.
- La aplicación debe entregar la duración del programa académico de cada país para así hacer una comparación a nivel de horas de trabajo, para ello se puede hacer un mapping directo, pero siempre pasando por el meta-modelo C.S.
- La aplicación debe entregar el “Awarding Body” que es el organismo a cargo de la certificación técnica no universitaria de cada país. Para obtener este tipo de información se puede realizar un mapping directo.

El siguiente listado describe los *Requerimientos No Funcionales* de la aplicación que de una u otra forma la delimitarán:

- El lenguaje utilizado para implementar el prototipo debe ser liviano, es decir, portable de manera que pueda ser instalado en cualquier estación de trabajo.
- El resultado obtenido al comparar competencias de una programa educacional con otro debe indicar a través de porcentaje la exactitud entre las competencias, donde el mejor match será dado por un 100%. Por lo tanto a partir de estos valores se pueda determinar qué competencias tienen un mejor match.
- La elección del motor de base de datos debe ser adecuado para aplicaciones Web donde la lectura de datos sea rápida y adecuada para el volumen de datos utilizado.

5.3 Diagramas de Casos de Uso

Se utilizarán diagramas de caso de uso para describir algunas de las funcionalidades de la aplicación que se implementará para mostrar los resultados obtenidos al utilizar las técnicas de Mapping y de Information Retrieval.

5.3.1 Diagrama de Caso de Uso: Generar C.S

Este caso de uso muestra los distintos tipos de información que pueden ser de ayuda para el experto al momento de comparar dos certificados de distintos países. Para ello primero se requiere que el especialista escoja el país y programa académico que desea comparar para así generar los respectivos C.S de cada país.

Como se muestra en el caso de uso de la figura 5.1 al generar el C.S se puede acceder a distinto tipo de información como se detalla a continuación:

- Determinar la duración del programa académico en el país de origen con la duración del programa dentro de su institución académica. Este tipo de información puede dar una idea al experto si los programas tienen una relación en cuando a horas de trabajo práctico y teórico.
- Conocer el “Awarding Body”, se requiere más bien para tener un conocimiento de cuál sería el homologo de este campo en cada país. Por ejemplo en el caso de Holanda el “Awarding Body” es “COLO”, por lo tanto, puede ser interesante para el especialista conocer cual institución dentro del framework italiano realiza las mismas funciones de coordinación desempeñadas por “COLO” o si se requiere más de una institución que supervise los programas académicos.
- Comparar la cantidad de competencias que tiene un país con respecto al otro y obtener el ranking de Similitud. Esta información será la que realmente permita poder determinar si dos programas académicos pueden ser homologables.
- Ver el Certificate Supplement generado para cada país con el objetivo de mostrar al experto la información en forma clara estándar.

La figura 5.1 muestra este caso de uso.

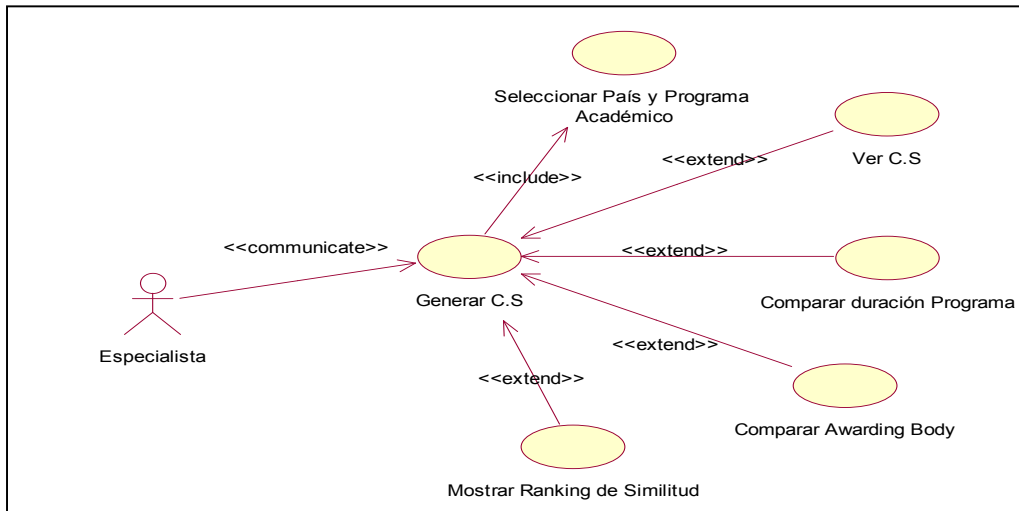


Figura 5.1. Diagrama Caso Uso: Generar C.S.

5.3.2 Diagrama de Caso de Uso: Mostrar Ranking de Similitud

Para realizar una comparación de competencias se requiere aplicar las técnicas de Information Retrieval que entregaran como resultado el ranking del grado de similitud de las competencias y habilidades de un perfil académico. Para ello es necesario que el certificado de cada país se encuentre en el formato del meta-modelo C.S.

El objetivo de la aplicación de estas técnicas a este nivel es ayudar al especialista a cargo de analizar los programas académicos en la decisión si un alumno cumple con las competencias o el grado de conocimientos suficientes para que sus estudios o parte de ellos sean homologados en el país de destino. La figura 5.2 muestra este diagrama de caso de uso.

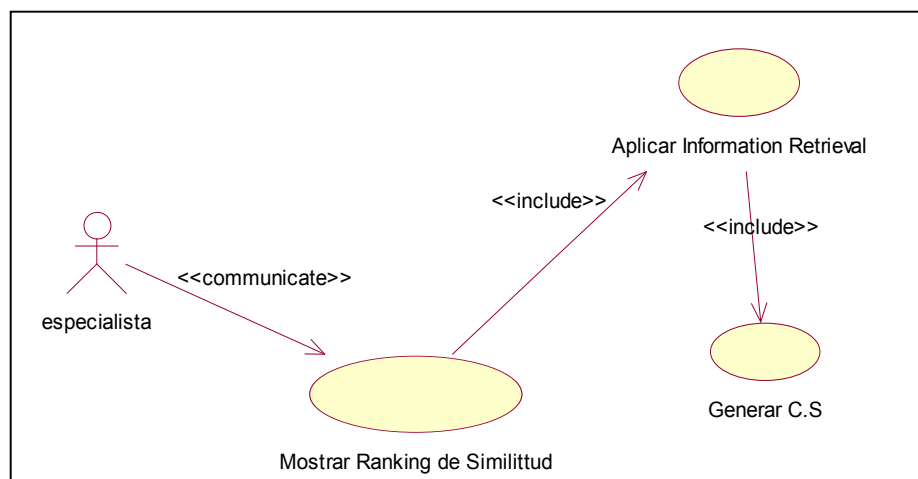


Figura 5.2. Diagrama Caso Uso: Mostrar Ranking de Similitud.

5.4 Diagramas de Secuencia

Los diagramas de secuencia que se mostrarán contienen detalles de la implementación de los escenarios en los cuales se utilizará la aplicación, incluyendo los objetos y clases que se usan para implementar estos escenarios, y mensajes pasados entre estos objetos.

5.4.1 Diagrama de Secuencia: Mostrar C.S

Este diagrama describe el escenario en el cual un experto desea ver el Certificate Supplement generado para cada país seleccionado. La figura 5.3 muestra el diagrama que detalla este escenario.

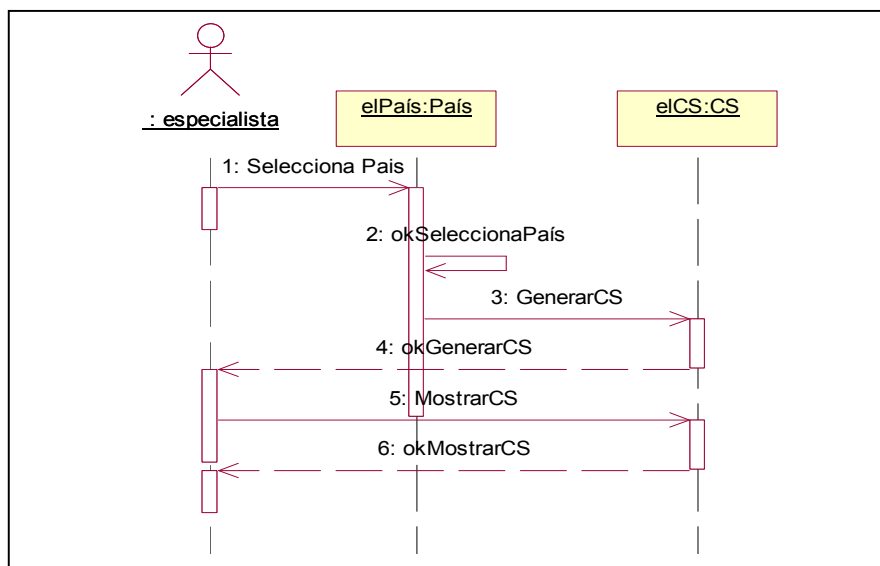


Figura 5.3. Diagrama de Secuencia: Mostrar C.S.

5.4.2 Diagrama de Secuencia: Mostrar Ranking de Similitud

Este diagrama describe el escenario en el cual un experto desea conocer cuál es la similitud entre las competencias de los programas académicos de una determinada pareja de países. La figura 5.4 muestra el diagrama que detalla este escenario.

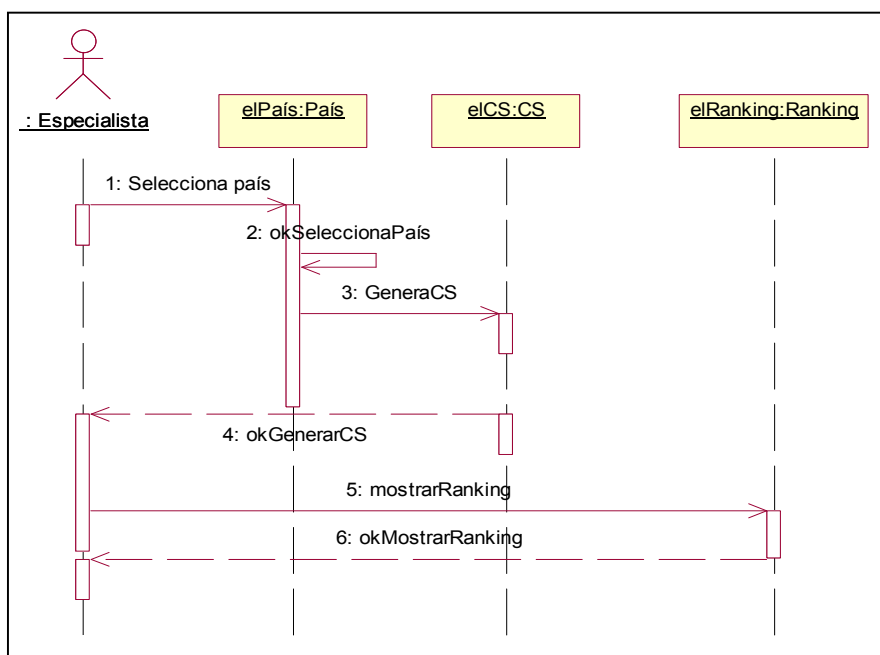


Figura 5.4. Diagrama de Secuencia: Mostrar Ranking de Similitud.

5.5 Modelos Entidad Relación

Con el conjunto de documentos de entrada que se recibió fue posible dar una formalización a los certificados de cada país que se tomaron como caso de estudio. Como se mencionó en el capítulo anterior cada uno de estos modelos fue formalizado a través de Redes Semánticas.

Con estos documentos ya formalizados, fue posible diseñar los modelos entidad relación a través de los cuales se creará la base de datos en MySQL, en la cual se almacenará la información necesaria para poder aplicar las técnicas de Information Retrieval y que sustentará la aplicación ilustrativa.

La decisión de utilizar este sistema de gestión de Base de Datos, se debe que es el más utilizado en aplicaciones WEB, sobre todo cuando se vincula a PHP, debido que es muy rápida en la lectura de los datos, sobre todo en el uso de aplicaciones WEB donde hay baja concurrencia en la modificación de los datos y en cambio el entorno es intensivo en la lectura de éstos. Por lo tanto, este sistema de base de datos es ideal para este proyecto debido que la aplicación de prueba será una aplicación WEB que se implementará utilizando el lenguaje PHP. Ya que el objetivo de la aplicación es mostrar la efectividad de los algoritmos de recuperación de información utilizados. La figura 5.5 muestra el modelo entidad relación de Italia y la figura 5.6 muestra el modelo entidad relación de Holanda.

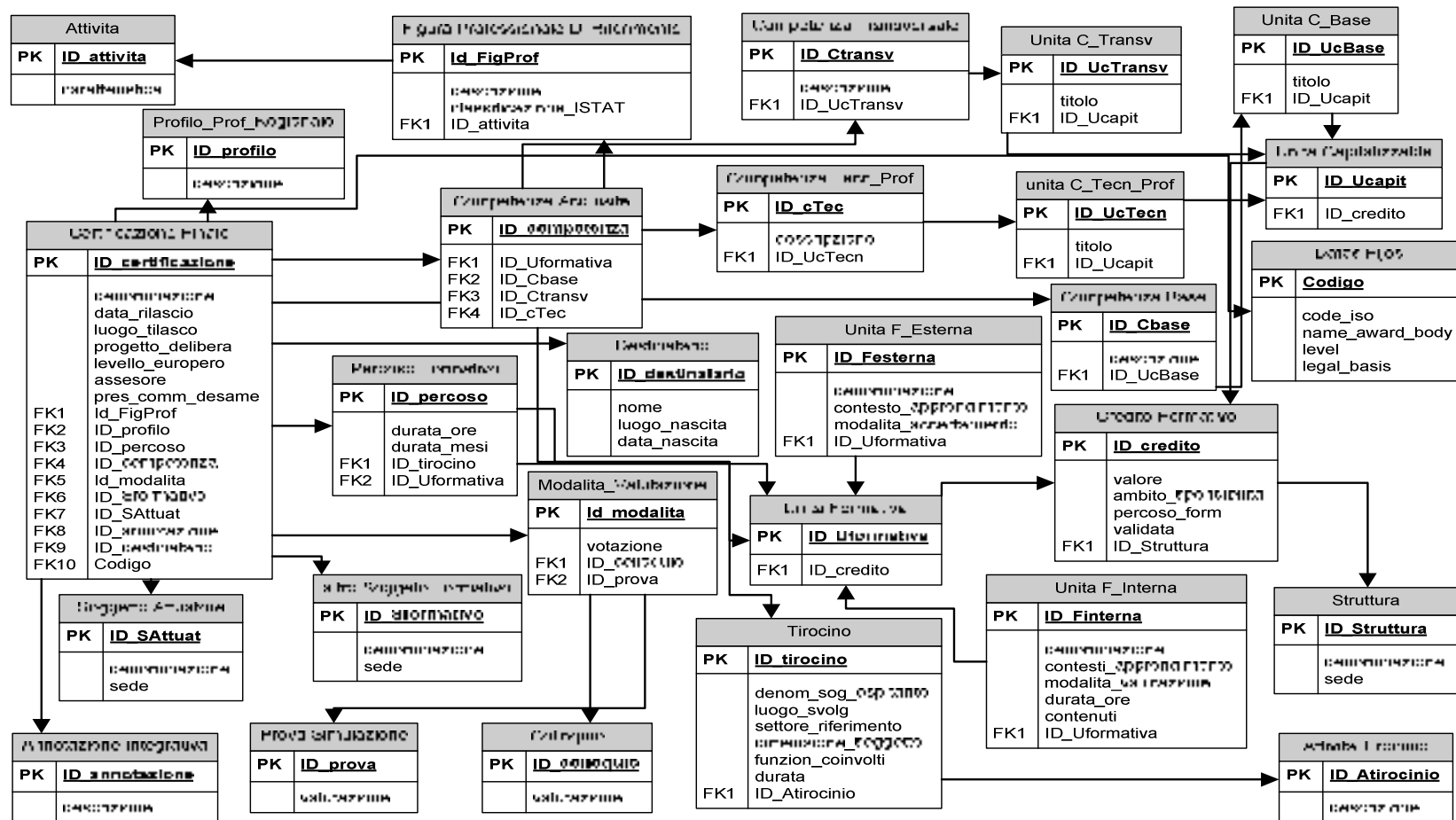


Figura 5.5. Modelo Entidad Relación Italia.

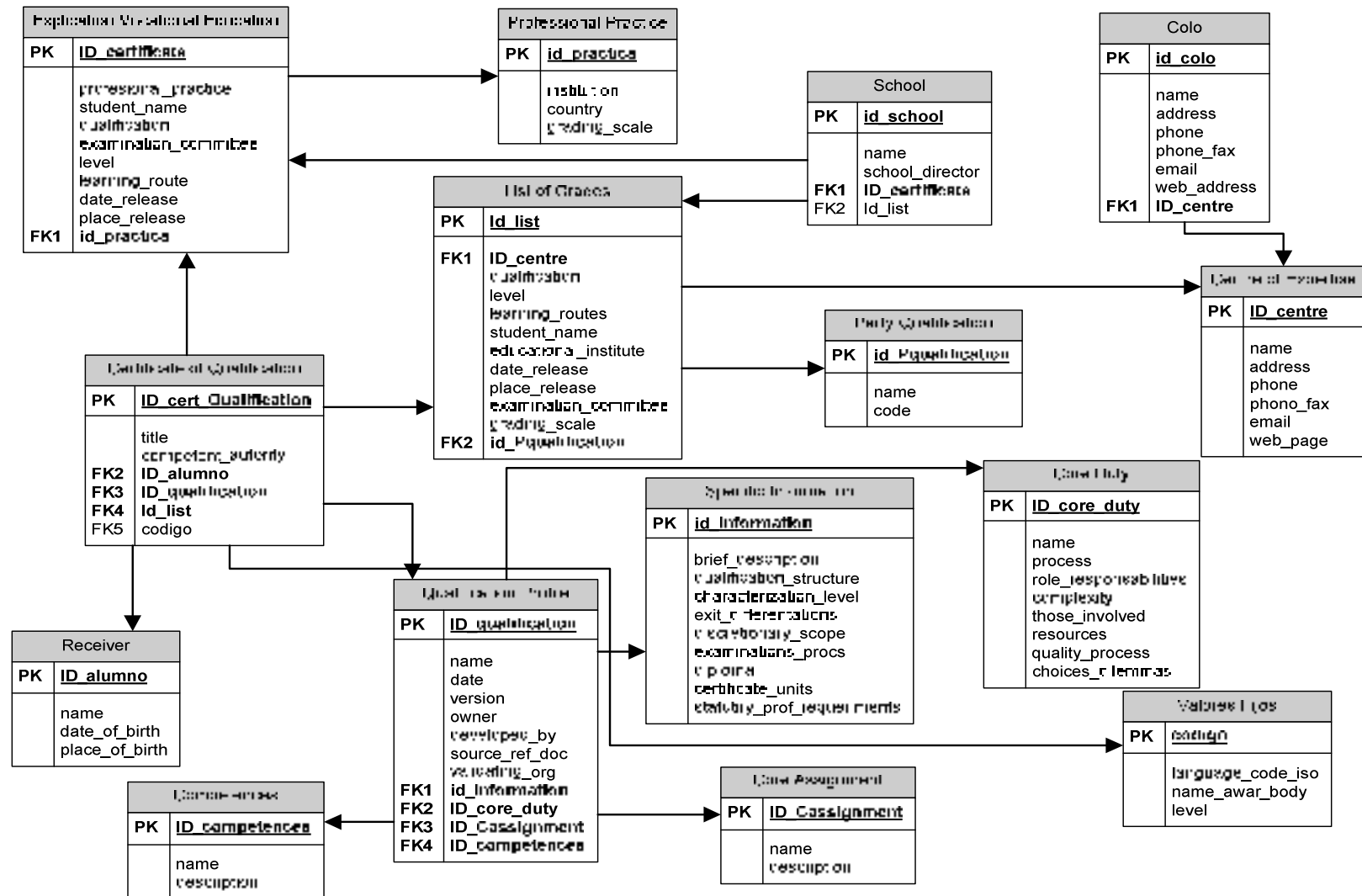


Figura 5.6. Modelo Entidad Relación Holanda.

5.6 Descripción Caso Estudio Holanda

El caso de estudio escogido corresponde a la carrera técnica no universitaria de “*Entrepreneur, Retail Trade*” (Empresarios, del comercio al detalle). A continuación se describe el perfil profesional de esta carrera.

Los empresarios del comercio al detalle trabajan como minoristas independientes utilizando una gama de fórmulas para el almacenamiento de productos alimenticios como no alimenticios.

Un profesional de esta área será capaz de manejar pequeños negocios autónomos, tener sus propias políticas para realizar sus planes de negocios que contendrán todas las ideas e información que consideren importante para su negocio. Además formularán sus propias políticas comerciales, financieras, organizacionales, personales y de logística en cuanto a la distribución de sus productos. Podrá tener personal a su cargo pero es él quien supervisa, anima, y motiva a su personal para que realice sus respectivas tareas de manera profesional. También mantienen una red de contactos para mantener su posición en el mercado.

Su enfoque dedicado al almacenamiento, la gama de productos, el contacto personal con sus clientes, logrará dar un servicio de excelencia y obtener la fidelización del cliente, con el objetivo de lograr amplios márgenes de ganancias. Otras de las tareas de los profesionales de esta área son fijar el curso que seguirá su compañía, por lo que combinan e idean procedimientos orientados a los resultados, para instalar nuevas sucursales, continuar o ampliar su negocio.

5.7 Descripción Caso de Estudio Italia

Este caso de estudio corresponde a la carrera de “*Tecnico Superiore Di Industrializzazione Del Prodotto e Del Processo*” (Técnico avanzado en la industrialización del producto y el proceso) un profesional de esta área debe ser capaz de analizar las instalaciones de la compañía según sus procesos de producción y costos relacionados. Analizando la estructura de ésta sobre las bases de sus líneas de negocios y organización económica. A la vez de estructurar planes de negocios óptimos de acuerdo al negocio y los propósitos estratégicos.

Los profesionales de esta área desarrollan un criterio de análisis que pueden aplicar a cada una de las funciones de la compañía y a los objetivos operacionales individuales de los principales procesos del negocio. Junto proporcionar soluciones útiles a los problemas organizacionales y económicos de la compañía dentro de los alcances operativos.

5.8 Descripción de Information Retrieval Aplicado a los Casos de Estudio

Las competencias y habilidades que describen a cada una de las carreras seleccionadas como caso de estudio se encuentran en tablas dentro de la base de datos. La base de datos generada se implementó con los modelos entidad relación mostrados en la sección anterior. Para el caso de Holanda la tabla que contiene dicha descripción es “ca_competencias_hol” y contiene 13 registros. Para el caso de Italia la tabla se denomina “ca_competencias_ita” y contiene 17 registros, en ambas tablas se almacena el nombre de la competencia y la descripción de ésta.

El primer paso para aplicar Information Retrieval es transformar toda la información que se describe en las tablas mencionadas en el párrafo anterior en una representación de vector.

Para ello se creó una tabla en la base de datos en la que cada uno de sus registro es una palabra que proviene de la unión de las tablas que describen las competencias de cada uno de los países, el nombre de la tabla es “ca_palabras”.

Para poder almacenar cada una de las palabras de los documentos, entiéndase como documento la descripción de cada una de las competencias de las tablas antes mencionadas, se implementó un pequeño programa en lenguaje C, que toma un archivo plano txt, del cual se han eliminado todos los signos de puntuación, generando un archivo plano txt en el que se imprime la instrucción que insertará cada una de las palabras en la base de datos, como muestra el ejemplo de la tabla 5.1. La cantidad total de registros que se generó al subir este archivo a la base de datos para cargar la tabla “ca_palabras” fue de 1061 registros inicialmente.

Tabla 5.1. Parte de archivo generado para llenar la tabla “ca_palabras”.

insert into ca_palabras values ('Entrepreneurs');
insert into ca_palabras values ('draw');
insert into ca_palabras values ('up');
insert into ca_palabras values ('a');
insert into ca_palabras values ('business');
insert into ca_palabras values ('plan');
insert into ca_palabras values ('incorporating');
insert into ca_palabras values ('the');

Dentro de la tabla se almacenaron todas las palabras que conforman la descripción de cada una de las competencias que detallan los distintos programas académicos incluyendo

aquellas que dan estructura al texto, es decir, conjunciones, preposiciones, artículos, pronombres, adjetivos, etc. Estas palabras se denominan “*Stop Words*” y no proporcionan ningún tipo de información, y por lo tanto no son necesarias y deben ser eliminadas para así disminuir el tamaño de la tabla que se utilizará posteriormente como vector cuando se aplique Information Retrieval.

Para eliminar todas aquellas palabras que no proporcionan ningún tipo de información se creó una tabla llamada “*ca_stopwords*” donde se almacenaron todas aquellas palabras que son consideradas Stop Words. Una vez que se consiguió tener una cantidad suficientes en esta tabla se realizó una consulta Sql que elimina de la tabla “*ca_palabras*” todos aquellos registros que coincidan con alguno de la tabla “*ca_stopwords*”. El resultado de la consulta produjo que la tabla “*ca_palabras*” disminuyera a 594 registros reduciendo el volumen de ésta en casi un 50%.

Como se describe, este proceso es relativamente fácil, lo complicado es la creación del listado de *Stop Words*, ya que se debe crear uno lo más global posible de manera de cubrir la mayor cantidad de palabras para poder obtener buenos resultados, por lo tanto, fue necesario hacer un análisis de los casos de estudio utilizados para determinar qué palabras serían consideradas *Stop Words* y cuáles no, este no fue un trabajo menor ya que el idioma de los documentos analizados se encuentra en Inglés y fue necesario realizar un minuciosos estudio. A continuación la tabla 5.2 muestra un pequeño listado de algunas de las palabras que componen la tabla “*ca_stopwords*”:

Tabla 5.2. Listado de algunos Stop Words.

i'm	Web	your's
Wouldn't	etc	he's
She's	It's	Isn't
And	Or	The
You're	We're	I'll
That	These	Tose

Antes de aplicar Information Retrieval y específicamente el algoritmo del Clasificador TFIDF fue necesario analizar otro aspecto importante que es hacer un análisis desde el punto de vista morfológico de cada una de las palabras donde muchas variantes pueden tener interpretaciones semánticas muy parecidas, para el caso de este proyecto la semántica de la palabra no es estrictamente necesaria ya que interesa únicamente su idea principal. Por ejemplo la palabra computador significa “*Ordenador, máquina electrónica que trata automáticamente la información y que ejecuta procesos lógicos a gran velocidad*”. Sí comparamos su significado con la palabra computadores, su plural se observa que intrínsecamente se está hablando del mismo concepto, porque cada uno de los

computadores es un “*Ordenador, máquina electrónica que trata automáticamente la información y que ejecuta procesos lógicos a gran velocidad*”. El plural en este caso como en muchos otros no resulta una característica muy interesante. Por lo tanto para evitar estas situaciones se aplicó la técnica de “*Stemming*” a la tabla “*ca_palabras*”.

La técnica de “*Stemming*” se trata de extraer la raíz o primitiva (Stem) que define una misma familia de palabras. De esta manera se llega a unificar y centralizar todo el texto en un conjunto de palabras básicas que facilitan considerablemente su búsqueda. Por ejemplo, dentro del caso de estudio se encontró con la siguiente familia de palabras: empresa, empresas, empresarial, empresario, empresarios, etc. En conclusión con la aplicación de esta técnica se minimizó considerablemente el diccionario de palabras del total del documento, ya que se ha conseguido sustituir la mayoría de éstas por la forma más primitiva del conjunto. A continuación la tabla 5.3 muestra un pequeño ejemplo del contenido de la tabla “*ca_stemming*”.

Tabla 5.3. Ejemplo del resultado obtenido luego de aplicar Stemming.

Lista de palabras antes de Stemming	Lista de palabras utilizando Stemming almacenadas en la tabla del mismo Nombre
Accept	Accept
Acceptable	
Accordance	Accord
According	Achiev
Achievement	
Achieving	

Realizar Stemming aporta dos ventajas. La primera es la reducción del almacenamiento de palabras tanto en memoria física como en memoria caché, habiendo un mejor aprovechamiento de los recursos. Como segunda ventaja está el incremento de velocidad y resultados en cada una de las consultas que se hagan a esta tabla puesto que se reducen en gran medida las posibles entradas.

Para este proyecto el uso de esta técnica es muy importante. Ya que al trabajar con un documento de entrada que contenía una cantidad muy elevada de datos es necesario minimizar lo más posible la información redundante, es decir, que den poco valor. De esta manera se gana en resultados y rapidez.

Encontrar el conjunto de palabras que pertenecen a una misma familia fue una tarea que requirió bastante tiempo ya que la documentación de entrada se encuentra en idioma inglés por lo tanto fue necesario hacer un minucioso análisis de está para poder determinar

qué palabras daban información redundante, para luego poder aplicar un conjunto de reglas que darían como resultado la raíz de la palabra.

Al aplicar Stemming los resultados obtenidos se almacenaron en una tabla denominada “*ca_steming*” en la cual se almacenó el stem de cada palabra para luego generar una consulta que devolviera el stem con la cantidad de veces que aparece dentro de la tabla, el resultado de esta consulta fue almacenado en una nueva tabla llamada “*ca_stem*”, en la cual se almacenaron aquellos stem donde el valor “*cantidad*” no fuera menor 3, este valor posteriormente se denomina peso de la palabra dentro del documento al aplicar I.R. Se deja sólo aquellas palabras con un peso mayor o igual a 3 porque en I.R se asume que palabras con un peso menor a este no aportaran información al momento de aplicar el algoritmo. Quedando así la tabla “*ca_stem*” constituida por 64 registros.

Con la tabla mencionada en el párrafo anterior, se pudo crear un programa en Lenguaje C que implementa el algoritmo del Clasificador TFIDF para poder implementar este algoritmo utilizó un vector de String correspondiente al vector de pesos, el cual contenía las palabras de la tabla “*Stem*” y tres matrices, cada una de éstas con dimensiones de 30x64, es decir, 29 filas que corresponden al documento que contiene las competencias italianas y holandesas y 63 columnas correspondientes al vector de stem. Las competencias italianas y holandesas son leídas desde un archivo plano llamado “*hol_ita_stem.txt*”. A este archivo también fue necesario aplicar la técnica de Stemming para poder realizar la comparación entre el vector de pesos y el documento y así obtener la frecuencia de cada palabra.

La primera matriz se utiliza para almacenar la cantidad de veces que una palabra del vector aparece en cada una de las filas del archivo “*hol_ita_stem.txt*”. Esto es necesario ya que estos valores se utilizan para poder obtener el cálculo de la fórmula 2.6 que se encuentra en el capítulo dos correspondiente a: $d^{(i)} = TF(w_i, d) \cdot IDF(w_i)$.

El número de documentos, que corresponde a cada una de las filas que describe las competencias, en el cual la palabra w_i aparece por lo menos una vez es almacenado en un arreglo denominado “*veces*”. Al tener estos cálculos se puede obtener cada uno de los $d^{(i)}$ de la fórmula 2.6, que son almacenados en la “*matriz2*”.

Finalmente en la tabla “*matriz3*” se almacena la similitud entre las palabras de las competencias italianas con las holandesas, para ello se dividió el archivo en dos partes, donde la primera corresponde a las competencias holandesas y la segunda a las italianas. La idea de realizar esto es generar cada uno de los valores correspondientes a la fórmula 2.10 del capítulo dos correspondiente a:

$$H_{TFIDF}(d') = \operatorname{argmax}_{c_j \in C} \frac{\sum_{i=1}^{|F|} c_j^{(i)} \cdot d'^{(i)}}{\sqrt{\sum_{i=1}^{|F|} (c_j^{(i)})^2}}$$

El numerador se obtiene de multiplicar cada uno de los valores almacenados en “matriz 2” de la siguiente manera, se toma cada valor correspondiente a cada una de las palabras de las competencias italianas y se multiplica por el valor de su correspondiente palabra holandesa, haciendo la sumatoria de los resultados. El denominador resulta de elevar al cuadrado cada uno de los valores de la “matriz2” e ir multiplicando las competencias italianas con las holandesas para realizar la posterior sumatoria de los resultados y finalmente sacar la raíz cuadrada del resultado final. Así al tener el cálculo del denominador y numerador se puede hacer la división final que nos dará como resultado el porcentaje de similitud entre las competencias italianas y holandesas.

5.9 Navegabilidad de la Aplicación

A continuación se muestran algunas pantallas de la aplicación implementada para así poder explicar de mejor manera la navegabilidad de ésta.

En la pantalla de inicio el especialista debe seleccionar los países y programas académicos que comparará posteriormente, para luego presionar el botón “Generar C.S” y obtener el Certificate Supplement de cada programa académico seleccionado. La figura 5.7 muestra esta pantalla.



Figura 5.7. Pantalla de Inicio

Una vez que el sistema genera los C.S pasa a la pantalla mostrada en la figura 5.8, donde puede seleccionar “Ver Certificate Supplement”, “Comparar Awarding Body”, “Comparar Programas” y “Mostrar Ranking de Similitud”. Para cualquiera de las opciones debe presionar el botón “Enviar”.



Figura 5.8. Selección de opciones.

Si escoge la opción “*Comparar Awarding Body*” mostrada en la figura 5.8, se desplegará una pantalla que muestra los “*awarding body*” de cada país, así el especialista puede ver cuál es la entidad homóloga en el país de destino que está a cargo del proceso de certificación. Como se muestra en la figura 5.9.

Italia	4. Official basis of the certificate: Awarding body, national	Holanda
Name: MIUR, Ministero dell'Università e della Ricerca Web: http://www.miur.it Address: Piazza J. F. Kennedy n. 20 00144 Roma EUR Tel: 06-97727124 Fax: E-mail: ufficiostampa@miur.it		COLO Agency Web: http://www.coloc.nl Address: Doos Pastoraalzaan 4 Tel: (079) 429 40 00 Fax: (079) 351 54 78 E-mail: info@coloc.nl

[VOLVER](#)

Figura 5.9. Comparar Awarding Body.

Si se escoge la opción “*Comparar duración del Programa*”, que se muestra en la figura 5.8, el experto podrá comparar la cantidad de tiempo dedicado al entrenamiento teórico y al entrenamiento práctico en cada país. Esta información puede ser útil para saber si al poseedor del certificado le falta completar alguno de estos tipos de entrenamiento dentro del país destino. La figura 5.10 ejemplifica esta situación.

Italia	6. OFFICIALLY RECOGNIZED WAYS OF ACQUIRING THE CERTIFICATE	Holanda
Total duration of the education / training leading to the certificate: Mode A Full time: 70 weeks Mode B Apprenticeship: 104 weeks		Total duration of the education / training leading to the certificate: Mode A Full time: 94 weeks Mode B Apprenticeship: 90 weeks

[VOLVER](#)

Figura 5.10. Comparar Duración del Programa.

Si se escoge la opción “Mostrar Ranking de Similitud” se mostrará una pantalla con las competencias de un determinado país. Para el caso de la figura 5.11 se muestran las competencias del programa de estudio italiano. Al seleccionar una competencia italiana se envía a una nueva pantalla donde se muestran todas las competencias holandesas que tienen un porcentaje de similitud con la competencia italiana seleccionada, como se muestra en la figura 5.12.



Figura 5.11. Selección de Competencia.



Figura 5.12. Mostrar Ranking de Similitud.

Si selecciona la opción “*Ver Descripción de Competencias*” que se muestra en la figura anterior se desplegará una pantalla que muestra el nombre de la competencia con la descripción de ésta y el porcentaje de similitud, como se indica en la figura 5.13.



ITALIA	DESCRIPCIÓN	SIMILITUD	HUNGARIA	DESCRIPCIÓN
Analizzare e pianificare le attività di sviluppo produttivo.	Analizzare il mercato, valutare le opportunità di sviluppo e pianificare le attività di sviluppo produttivo. Analizzare le opportunità occupazionali, valutare le risorse umane, le risorse finanziarie e le risorse materiali, organizzare e gestire le attività di sviluppo produttivo. Pianificare le attività di sviluppo produttivo, valutare le opportunità di sviluppo produttivo e pianificare le attività di sviluppo produttivo. Pianificare le attività di sviluppo produttivo, valutare le opportunità di sviluppo produttivo e pianificare le attività di sviluppo produttivo.	100.000 100 100 100 100 100 100	1. Termelési tevékenység 2. Közvetlen vezetői tevékenység 3. Vezetéktanulmányok 4. Termelési tevékenység 5. Szervezési tevékenység 6. Vezetéktanulmányok 7. Szervezés	1. Termelési tevékenység 2. Közvetlen vezetői tevékenység 3. Vezetéktanulmányok 4. Termelési tevékenység 5. Szervezési tevékenység 6. Vezetéktanulmányok 7. Szervezés

Figura 5.13. Mostrar Descripción de Competencias.

Si selecciona la opción “*Ver Certificate Supplement*” que se muestra en la figura 5.8 mostrará una pantalla como la de la figura 5.14, en la cual el especialista deberá escoger cual de los dos certificados quiere ver. Al presionar cualquiera de las opciones se desplegará el certificado correspondiente al país seleccionado como muestra la figura 5.15.



Figura 5.14. Selección de Certificados.

CERTIFICATE SUPPLEMENT

1. TITLE OF THE CERTIFICATE
ENTREPRENEUR RETAIL TRADE

2. TRANSLATE TITLE OF THE CERTIFICATE
ENTREPRENEUR RETAIL TRADE

3. PROFILE OF SKILLS AND COMPETENCES
A typical holder of the certificate is able to:

- the purchaser drawing a plan to give an impression of the company's objectives. While doing so they can draw attention to the financial and human resources.
- the purchaser make an inventory of sources for their business plan.
- the purchaser make use of a variety of sources of information for their business plan. The holder of the plan carries a copy outside of the business to bring to the company together with the financial and management plan.
- the purchaser draw up a cost calculation during the preparation of their business plan.
- the holder of the business plan to draw up a business plan in the form of a plan in terms of policy plan, taking into account the marketing plan, the human resource plan, the financial plan, and the personal policy plan.
- the purchaser make a marketing plan in which the policy is being to be produced, such as: supply, production, personal and financial.
- the entrepreneur to estimate the number of employees to be recruited, and the marketing plan, especially how this is necessary and/or to be recruited.
- The entrepreneur also draws up a logistics plan specifying the logistics process from purchase to retail sale.

4. RANGE OF OCCUPATIONS ACCESSIBLE TO THE HOLDER OF THE CERTIFICATE
Entrepreneurs Enterprise Retail

5. OFFICIAL BASIS OF THE CERTIFICATE

<p>Name and status of the body awarding the certificate</p> <p>COLO Agency web: http://www.colo.nl</p> <p>Boris Pasternaklaan 4 2719 DA Zoetermeer or Postbus 7259 701 AG Zoetermeer, Netherland Tel: 079 329 40 00 Fax: 079 351 54 78 E-mail: voorkichting@colo.nl</p>	<p>Name and status of the national / regional authority providing accreditation / recognition of the certificate.</p> <p>COLO Agency Tel: 079 329 40 00 Fax: 079 351 54 78 E-mail: voorkichting@colo.nl</p>
---	--

<p>Level of the certificate (national or international)</p> <p>Dutch National Framework of Qualification Level 4</p>	<p>Grading Scale / Pass Requirements</p> <p>10 excellent 9 very good 8 good 7 amply sufficient 6 sufficient 5 almost sufficient 4 insufficient 3 very insufficient 2 poor 1 very poor</p>
---	--

6. OFFICIALLY RECOGNIZED WAYS OF ACQUIRING THE CERTIFICATE

Description of Vocational Education and Training Received	Percentage of total programme (%)		Duration (hours / week / month / years)	
	Mode A	Mode B Full time Apprenticeship	Mode A	Mode B
School / Training centre based	57%	32%	54 weeks	29 weeks
Workplace based	43%	68%	40 weeks	61 weeks
Accredited prior learning	Available		94 weeks	90 weeks
Total duration of the education / training leading to the certificate			94 weeks	90 weeks

Figura 5.15. Certificate Supplement Holanda.

En este punto se detalló la navegabilidad de la aplicación implementada ayudándose de la utilización de las interfaces de ésta.

5.10 Resultados y Trabajos Futuros

La utilización del algoritmo del clasificador TFIDF aplicado a los casos de estudio descritos en los puntos anteriores de este capítulo dio unos porcentajes de similitud bastante bajos, todos menores a un 10%, donde el porcentaje mayor se dio con un 0.014% como se muestra en la figura 5.16, es decir la competencia italiana “*Product Engineering*” del conjunto de competencias holandesas con las que obtuvo un matchig más alto se dio con la competencia “*Goal and Procedure Oriented Actions*”.

ITALIA	SIMILITUD	HOLANDA
Product Engineering	0.014112	Goal and procedure oriented actions
Via Descripción Competencias	0.013989	Customer-oriented actions
	0.0139	Entrepreneurial actions
	0.012989	Professional actions
	0.01232	Orientation towards organization & surroundings
	0.01	Transnational actions

VOLVER

Figura 5.16. Competencias con mayor porcentaje de Similitud.

Al analizar la descripción de estas competencias se pudo notar que de una manera bastante global lo que se intenta es que el alumno logre desarrollar tareas según acuerdos y procedimientos dados. Pero en sí la descripción de la competencia italiana es mucha más específica y orientada a un tipo de tarea más industrial no así la descripción de la competencias holandesa que es más global y específica al rubro del comercio al detalle, como se muestra en la figura 5.17, la tabla 5.4 muestra la traducción de la descripción de las competencias con los match más altos.

ITALIA 	DESCRIPCIÓN	SIMILITUD	HOLANDA 	DESCRIPCIÓN
Product Engineering.	Analyse and interpret the data and the specifications of the product; choose the suitable production technologies on the criteria basis of production cost and expected quality; plan the working phases, according to procedures and standards and with precise reference to materials, components, production means, equipment, technological parameters, timing, etc.	0.014142 0.013989 0.0139 0.012989 0.01232 0.01	1. Goal- and procedure-oriented actions 2. Customer-oriented actions 3. Entrepreneurial actions 4. Professional actions 5. Orientation towards organization & surroundings 6. Entrepreneurial actions	1. Entrepreneurs in the retail trade are able to perform their duties in an adequate manner, and in accordance with the agreements and/or procedures. 2. Entrepreneurs in the retail trade are able to ensure for an adequate organization of the work that takes account of the wishes of others and ultimately ensures for a sustainable relationship. 3. Entrepreneurs in the retail trade are able to make the personal contribution required to identify opportunities and problems, and to take carefully-considered risks beneficial to the company. 4. Entrepreneurs in the retail trade are able to act in an appropriate manner on the basis of specific professional and/or sectoral knowledge and skills 5. Entrepreneurs in the retail trade are able to ensure for an adequate compatibility of the work with the organization's objectives and culture, and to implement developments in society and the profession in their work. 6. Entrepreneurs in the retail trade are able to make the personal contribution required to identify opportunities and problems, and to take carefully-considered risks beneficial to the company.

Figura 5.17. Comparación de descripción de Competencias con mayor porcentaje de Similitud.

Tabla 5.4. Descripción y Comparación de Competencias con Mayor Valor de Similitud.

Descripción Competencia Italiana	Descripción Competencia Holandesa
<p>Ingeniería de producto:</p> <p>Analizar e interpretar los datos y las especificaciones de los productos; seleccionar las tecnologías de producción más adecuada según los criterios de costos de producción y calidad esperada, planear las fases de trabajo de acuerdo con los procedimientos y estándares y con la referencia precisa de los materiales, componentes, medios de producción, equipos, parámetros tecnológicos, calendario, etc.</p>	<p>Acciones orientadas a los objetivos y procedimientos:</p> <p>Los empresarios de comercio son capaces de desempeñar sus funciones de manera adecuada, según los acuerdos y/o procedimientos.</p> <p>Acciones orientadas al cliente:</p> <p>Los empresarios de comercio son capaces de garantizar una adecuada organización del trabajo considerando los deseos</p>

	<p>de los demás, en definitiva asegurando una buena relación.</p> <p>Acciones empresariales:</p> <p>Los empresarios de comercio son capaces de realizar contribuciones personales para identificar oportunidades y problemas, considerando cuidadosamente los riesgos y beneficios para la empresa.</p>
--	--

El porcentaje más bajo obtenido corresponde a 0.01% lo que indica que la similitud entre las competencias es casi nula como se muestra en la figura 5.18. La figura 5.19 muestra la descripción de estas competencias. En la tabla 5.5 se traducen las competencias mostradas en la figura 5.19, con ello se puede ver que no hay relación entre las competencias italianas y holandesas.



Figura 5.18. Comparación de Competencias con menor porcentaje de Similitud.

ITALIA	DESCRIPTION	SIMILITUDE	HOLLAND	DESCRIPTION
Wood Technology	<p>Valutare le tecnologie e le materie prime dei materiali di prima mano e le tecnologie produttive, implementando i aspetti di base, come materiali e le tecnologie di controllo della qualità, e le competenze trasversali, come la capacità di costruire un team, la capacità di risolvere i problemi e le tecnologie usate nella wood and furniture industry.</p>	<p>0.01 0.01 0.01 0.01</p>	<p>2 Goals and procedures oriented action 3 Business orientation 7 Customer orientation 4 Ability to solve problems and situations</p>	<p>1. Entrepreneurial skills and methods able to perform the business activities and procedures. 2. Ability to solve the business problems and to make the process decisions required to find the quality and problems and to identify the materials and the methods for company. 3. Ability to solve the business problems, determine the organization of the work that takes into account the interests of all the stakeholders, with the potential relationship. 4. Ability to solve the business problems, adopt an appropriate approach to the thinking, analysis and solution of problems, the communication, the control and cooperation of resources in the form of specifications.</p>

Figura 5.19. Descripción de Competencias con menor porcentaje de Similitud.

Tabla 5.5. Descripción y Comparación de Competencias con Menor Valor de Similitud.

Descripción Competencia Italiana	Descripción Competencia Holandesa
<p>Tecnología de la madera:</p> <p>Evaluar el valor tecnológico y el uso de las materias primas poniendo atención a los aspectos histológicos, fisiológicos y morfológicos de las principales maderas; realizando pruebas sobre las características reológicas de las principales maderas y de la composición de materiales, analizando sobre una base científica, los mecanismos que regulan los procesos industriales y las tecnologías usadas en la madera y en la industria de muebles.</p>	<p>Acciones orientadas a los objetivos y procedimientos:</p> <p>Los empresarios de comercio son capaces de desempeñar sus funciones de manera adecuada, según los acuerdos y/o procedimientos.</p> <p>Acciones empresariales:</p> <p>Los empresarios de comercio son capaces de realizar contribuciones personales para identificar oportunidades y problemas, considerando cuidadosamente los riesgos y beneficios para la empresa.</p> <p>Acciones orientadas al cliente:</p> <p>Los empresarios de comercio son capaces de garantizar una adecuada organización del trabajo considerando los deseos de los demás, en definitiva asegurando una buena relación.</p> <p>Habilidad para tomar decisiones y resolver problemas:</p>

	<p>Los empresarios del comercio están en condiciones de adoptar enfoques para la identificación, análisis y evaluación de problemas, tomado decisiones relativas a las opciones y a la interpretación de conclusiones en términos de las acciones específicas.</p>
--	--

Según estos resultados obtenidos se puede decir que si bien la aplicación de la técnica no dio los resultados esperados debido principalmente a la orientación profesional de cada carrera, que es totalmente distinta ya que el perfil holandés entrega una base profesional dirigida al comercio al detalle. Mientras que el perfil italiano está orientado a la formación de profesionales que se dedicarán a trabajar en los procesos productivos de un determinado producto dentro del área industrial. Por lo tanto, en si los perfiles no tienen ninguna relación. Pero a pesar de esto se puede decir que la técnica de Information Retrieval sí sirve para poder comparar competencias o hacer comparaciones semánticas ya que con los resultados obtenidos y el análisis de éstos se pudo determinar que los planes académicos no tiene ninguna similitud, por consiguiente, si un estudiante con el título profesional holandés de *“Entrepreneur, Retail Trade”* quisiera que sus conocimientos fueran convalidados en la carrera italiana de *“Tecnico Superiore Di Industrializzazione Del Prodotto e Del Processo”* no se podría ya que no son compatibles ambas carreras.

Como trabajo futuro se podría buscar otra técnica dentro de lo que exista en *“Matching Learning”*, *“Técnicas de Aprendizaje”* o WordNet que se encuentran dentro de la categoría de recuperación de información para así poder hacer una comparación y determinar si estas técnicas aportan una mejora a la utilizada acá obteniéndose mejores resultados y puedan complementarla. Ya que la técnica utilizada en este trabajo es más bien básica y experimental en esta área de estudio que corresponde a la comparación semántica de competencias académicas.

Capítulo 6

Conclusión

La primera etapa de este proyecto se basó en la adquisición de conocimientos sobre los conceptos involucrados en la temática del proyecto. Con el estudio de publicaciones y trabajos existentes al respecto se pudo lograr una visión global de las diferentes áreas que hay en informática para integrar modelos semánticamente diferentes.

La primera técnica de integración de documentos analizada fue el Mapping de Ontologías, de los procesos de mapping se puede destacar que éstos ayudan a estandarizar la descripción de la información.

Otra área de estudios fue Knowledge Sharing que mostró la utilización de distintas técnicas para comunicar modelos, poder compartir el conocimiento y reutilizarlo, la mayoría de estos trabajos destacan la utilización de mediadores y reglas basadas en éstos que resuelven problemas de heterogeneidad semántica realizando una integración flexible de varios sistemas de información.

La última área analizada fue Information Retrieval donde se vieron distintas técnicas de categorización de texto que se apoyaban de clasificadores probabilísticos, donde el que más destacaba como un posible aporte para el ámbito y envergadura de este proyecto fue el clasificador TFIDF que entrega la frecuencia de un término dentro de un conjunto de documentos.

Del estudio realizado en cuanto al estado del arte con respecto a integración de sistemas se decidió utilizar Information Retrieval como una metodología para poder comparar modelos semánticamente distintos, ya que los algoritmos utilizados por esta técnica dan como resultado la ocurrencia de un término dentro de un conjunto de documentos de entrada, lo que es útil para el enfoque de este trabajo que era lograr obtener el grado de similitud entre las instancias de competencias de dos modelos semánticamente distintos.

También como primera parte del trabajo se analizaron los procesos de certificación existentes en el ámbito europeo, pudiendo así reconocerse las necesidades no cubiertas por estos programas y que motivaron el desarrollo de este proyecto.

Otro aspecto importante de destacar es el análisis de las diferentes alternativas de formalización, dentro de las que se destacó Redes Semánticas por ser el mecanismo que más se adecuaba al perfil de las personas que utilizarán estos modelos y que no están necesariamente relacionadas con el área informática.

La segunda parte de este proyecto constó de la selección de países que se utilizarían como casos de estudio, la decisión tomada dentro de esta parte fue clave debido que con ellos se comprobaría la eficiencia de la técnica de Information Retrieval. Esta técnica específicamente se implementó para la parte de competencias y habilidades que se detallan dentro del Certificate Supplement de cada país porque la información dentro de este punto no es uniforme, y por lo tanto se hace necesario obtener un grado de similitud lo más exacto posible para poder determinar si los conocimientos adquiridos por un estudiante corresponden al perfil educacional de la institución en la cual éste desea completar o complementar sus estudios.

También la identificación de requerimientos funcionales como no funcionales cobró vital importancia en esta etapa del trabajo, porque una vez que estos estuvieron claramente definidos se pudo comenzar a implementar la aplicación ilustrativa que sirvió como interfaz para mostrar los resultados de aplicar la técnica de integración escogida.

El mayor desafío fue finalmente la implementación de Information Retrieval, específicamente la implementación del algoritmo del Clasificador TFIDF, ya que éste entrega la frecuencia de un término dentro de un conjunto de documentos, pudiéndose así obtener la similitud entre competencias.

Al aplicar la técnica antes mencionada en los casos de estudio seleccionados se obtuvo que ésta entrega bajos resultados de similitud, todos menores a un 10%. La información obtenida se complementó analizando las competencias con mayor y menor valor de similitud, a través de una traducción, lectura y comparación de éstas. Que determinó que los planes de estudios no pueden ser considerados similares debido que el perfil holandés entrega una base profesional dirigida al comercio al detalle. Mientras que el perfil italiano está orientado a la formación de profesionales dedicados a los procesos productivos de un determinado producto dentro del área industrial.

Por lo expuesto en el párrafo anterior se puede determinar que los perfiles académicos no tienen ninguna relación. Pero a pesar de esto la técnica de Information Retrieval sí sirve para poder comparar competencias o hacer comparaciones semánticas ya que con los resultados obtenidos y el análisis de éstos se pudo determinar que los planes académicos no tienen ninguna similitud cumpliéndose así el objetivo principal del proyecto que era desarrollar un modelo conceptual de integración basado en el Certificate Supplement que

permitiera comparar certificados de estudio para algunos países de la Unión Europea, semánticamente distintos.

Del trabajo realizado queda la satisfacción de haber desarrollado un tema completamente nuevo del cual no se tenían antecedentes anteriores. Como trabajo futuro se propone mejorar la técnica para que los resultados devueltos sean aún más exactos o quizás buscar una nueva técnica dentro de lo que exista en “*Matching Learning*”, “*Técnicas de Aprendizaje*” o “*WorNet*” de manera que se automatice el proceso de cálculo y de obtención de palabras como los Stopwords y puedan así quizás complementar la técnica utilizada en este proyecto.

Referencias

- [1] Building European Passport for Transparent and International Certification.
<http://www.icc-languages.eu/beatric.php>
- [2] European Commission, Europass.
http://www.mec.es/europass/html/que_es.html
- [3] Noticias Jurídicas, Base de datos de legislación tratado 149.
http://www.juridicas.com/base_datos/Admin/ttce.13t11.html.
- [4] National VET Systems, Thematic Overviews.
http://www.trainingvillage.gr/etv/Information_resources/NationalVet/Thematic/
- [5] European Commission, Sistema Europeo de Transferencia y acumulación de Créditos (ECTS).
http://europa.eu.int/comm/education/programmes/socrates/ects/index_es.html
- [6] European Commission, European Credits Transfert System for Vocational Education and Training (ECVET).
http://www.embedding-standards.com/ecvet_002.pdf
- [7] National Qualifications Authority of Ireland.
<http://www.nqai.ie/>
- [8] Towards a European Qualifications Framework: A View from the European Consortium for Accreditation (ECA) February 2004.
http://www.bologna-bergen2005.no/b/BFUG_Meetings/040309Dublin/0402ECA_on_EQF.pdf
- [9] European Commission, European Qualification Framework.
http://ec.europa.eu/education/policies/educ/eqf/index_en.html
- [10] European Commission, Socrates- Erasmus.
http://ec.europa.eu/education/programmes/llp/erasmus/index_en.html
- [11] European Commission, Leonardo da Vinci.
http://ec.europa.eu/education/programmes/llp/leonardo/index_en.html
- [12] Thomas R. Gruber (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing.

- [13] H.Wache, T. Vögele, U.visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S (2001). Hübner, Ontology- Based integration of Information – A Survey of Existing Approaches.
- [14] Floris Wiesman, Nico Roos (2004).Domain Independent of Ontology Mapping.
- [15] Nicholas J. Belkin and W.Bruce Croft, Information filtering and information retrieval: Two sides of the same coin?, Communication of the ACM, Diciembre 1992, v35 n12 p29 (10).
- [16] Thorsten Joachims (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features.
- [17] Thorsten Joachims (1997). A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization.
- [18] David D. Lewis (1991). Learning in Intelligent Information Retrieval.
- [19] Christoph Goller, Joachim Löning, Thilo Will, Werner Wolff. Automatic Document Classification: A thorough Evaluation of various Methods.
- [20] Mike Uschold, Rob Jasper, Peter Clark (1999). Three Approaches for Knowledge Sharing: A comparative Analysis.
- [21] Mike Uschold, Rob jasper (1999). A Framework for Understanding and Classifying Ontology Applications.
- [22] Holger Wache (1999). Towards Rule-Based Context Transformation in Mediators.
- [23] European Commission, Centro Nacional Europass, Curriculum Vitae.
<http://www.mec.es/europass/html/curriculum.html>
- [24] European Commission, Centro Nacional Europass, Pasaporte de Lenguas Europass.
<http://www.mec.es/europass/html/pasaporte.html>
- [25] European Commission, Centro Nacional Europass, Documento de Movilidad Europass.
<http://www.mec.es/europass/html/movilidad.html>
- [26] European Commission, Centro Nacional Europass, Certificate Supplement Europass.
http://www.mec.es/europass/html/suplemento_certificado.html

[27] European Commission, Centro Nacional Europass, Suplemento Europass al Título Superior.

http://www.mec.es/europass/html/suplemento_titulo.html

[28] B.E.A.T.R.I.C Project, Formalization of Conceptual Elements and Descriptive Ítems, Noviembre 2005.

[29] Further Education and Training council Awards.

<http://www.fetac.ie/>