

Judit Feliu. *Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica*. Tesis doctoral. Orientadora: Maria Teresa Cabré. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 2004.

Judit Feliu (Universitat Pompeu Fabra, España)

El objetivo principal de esta tesis doctoral es establecer las bases para un sistema de detección semiautomática de relaciones conceptuales a partir de textos especializados. Para alcanzar este objetivo, se presenta una definición de relación conceptual en el marco de un enfoque comunicativo de la terminología y, a partir de esta definición, la autora detecta y aplica diversos marcadores lingüísticos verbales para aislar fragmentos textuales que contengan unidades de conocimiento especializado, vehiculadas mediante términos y relaciones conceptuales en el ámbito del genoma humano.

Para llegar a proponer un prototipo de sistema de detección semiautomática de relaciones conceptuales se siguen diversas etapas desde el punto de vista teórico y aplicado. Los objetivos específicos de la tesis son, en primer lugar, validar empíricamente la tipología de relaciones conceptuales establecida en un trabajo de investigación previo. En este sentido, se han aplicado todos los marcadores lingüísticos explícitos que vehiculan cada tipo de relación conceptual sobre un corpus de genoma humano y se ha actualizado la lista de marcadores de relación conceptual. La lista definitiva resultante en este trabajo contiene 55 unidades verbales (acompañadas, a menudo, de preposiciones y elementos nominales de soporte). En segundo lugar, se establecen las estrategias sintáctico-semánticas para detectar con el máximo de precisión los marcadores que expresan cada tipo de relación. En este sentido, además de fijar el patrón sintáctico que expresa la relación, se aplica la estrategia del factor del contexto, una estrategia semántica, que permite determinar con un grado de fiabilidad bastante elevado si un marcador expresa un tipo de relación u otro. Esta estrategia semántica se basa en la validación del vínculo de un candidato a término con una ontología del ámbito de especialidad. El resultado de la aplicación de estrategias sintáctico-semánticas es la obtención de fragmentos textuales especializados que contienen, como mínimo, dos unidades terminológicas vinculadas entre sí mediante un tipo de relación conceptual.

El contenido de este trabajo se estructura en capítulos que siguen el orden de los objetivos específicos anteriores. El capítulo 2 presenta los resultados más destacados de la tesina de doctorado. Se retoma la noción de relación conceptual y se presenta la tipología de relaciones conceptuales indicando las principales características de cada tipo de relación.

El capítulo 3, y con el propósito de demostrar el vínculo estrecho entre léxico y ontología, se detalla el tratamiento de las relaciones conceptuales en las principales ontologías utilizadas para la recuperación de información. Se observa que, tradicionalmente, no se ha distinguido entre la noción de relación y la unidad léxica que expresa esta relación y, además, se introduce y se valida la tipología de relaciones conceptuales en una ontología sobre el ámbito del genoma humano que forma parte de un banco de conocimiento sobre este ámbito especializado.

En el capítulo 4 se presenta el proceso de constitución del corpus de trabajo. Para cada tipo de relación, se detallan los posibles marcadores lingüísticos aplicados sobre el material textual y se indican los resultados obtenidos cuanto a precisión y ruido. Seguidamente, se muestran las estructuras más significativas para cada tipo de relación. Estas estructuras devienen el punto de partida para la propuesta de sistema de detección semiautomática.

En el marco del capítulo 5, se describen detalladamente las estrategias que pueden refinar el sistema de detección semiautomática de relaciones conceptuales junto con la extracción de las unidades candidatas a ser términos. La combinación de estas dos herramientas orientadas a detectar términos y relaciones proporciona un primer esqueleto de la información especializada de los textos.

El capítulo 6 contiene la descripción del sistema de detección de relaciones conceptuales de manera sistematizada y próxima a su futura implementación. Finalmente, el capítulo 7 contiene las conclusiones de la investigación y el capítulo final las referencias bibliográficas consultadas. Se incluye un CD-ROM anexo que contiene el corpus de análisis y la base de datos de marcadores verbales de relación conceptual con un total de 3.114 contextos etiquetados.

Las aportaciones principales de esta investigación son las siguientes:

- a) Se ha fijado una definición de la noción de relación conceptual sobre la base de un trabajo de corpus especializado y la detección de unidades terminológicas *in vivo* que vehiculan los conceptos mediante una relación. La tipología de relaciones conceptuales supera la extensa descripción tradicional de las relaciones jerárquicas y presenta un catálogo más amplio.
- b) Se ha establecido una lista de marcadores verbales que vehiculan una o más relaciones conceptuales. Estos marcadores son el punto de partida del programa de detección terminológica *Mercedes* para detectar, en un primer nivel, los contextos que contienen unidades terminológicas pero también posibles materializaciones de relaciones conceptuales.

- c) Se ha integrado la tipología de relaciones conceptuales en una ontología sobre el genoma humano y las unidades terminológicas que se entran en la base de datos terminológica ligada a la ontología están, asimismo, vinculadas con otros conceptos gracias a una o diversas relaciones conceptuales.
- d) Se han utilizado herramientas automáticas de detección y extracción de términos (*Mercedes* y *Yate*) para intentar evitar al máximo la detección manual. En este sentido, la combinación de las herramientas y la descripción de nuevas estrategias lingüísticas para retener o no un marcador verbal serán el punto de partida para la futura implementación del prototipo de sistema de detección semiautomática de relaciones conceptuales que servirá, además, para refinar el sistema de detección de candidatos a términos.

Finalmente, la tesis presenta diversas futuras vías de investigación que se abren ante los resultados obtenidos. Destaca, de entre las posibilidades de continuación de la investigación, la implementación y combinación del sistema de detección semiautomática de relaciones conceptuales y el programa de extracción de candidatos a términos para obtener nudos de conocimiento especializado que alimenten el sistema de representación tridimensional del contenido especializado de un texto mediante la visualización de nudos y relaciones.