

TERMINUS, GESTIÓN DE CORPUS Y TERMINOLOGÍA EN LÍNEA

M. Amor Montané March (Grupo IULATERM – Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra)

TERMINUS es un sistema de gestión de corpus y terminología desarrollado por el grupo IULATERM, grupo de investigación en léxico, terminología, discurso especializado e ingeniería lingüística del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra de Barcelona. Ha sido diseñado e implementado por M. Teresa Cabré, Gabriel Reus y Rogelio Nazar, con la colaboración de Iria da Cunha y M. Amor Montané.

The screenshot shows the TERMINUS web interface. At the top, there is a header with the logo and the text "Gestió de corpus i terminologia en línia". Below the header, there is a login section with fields for "Lengua:" (set to "Castellano"), "Nombre de acceso:", and "Contraseña:". There is a link for "¿Ha olvidado la contraseña?" and an "Aceptar" button. Below the login section, there is a "Versión demo" link and "Instrucciones de uso:" with links for "catalán" and "castellano", and a "Créditos" link. To the right of the login section, there is a description of TERMINUS: "TERMINUS es una estación de trabajo para la terminología. Integra la gestión de corpus y de terminología. Permite crear y gestionar grupos de trabajo y modelar las categorías de datos. Incluye la cadena completa del trabajo terminográfico individual y en equipo: búsqueda, constitución y exploración de corpus textuales, gestión de glosarios y proyectos, creación y mantenimiento de bases de datos y edición de diccionarios." Below this, it says "Terminus consta de diversos módulos articulados:" followed by a list of modules: "Proyectos", "Fuentes", "Estructuración conceptual", "Documentos", "Corpus", "Análisis", "Glosarios", and "Términos", each with a brief description of its function. At the bottom, there are logos for "UNIVERSITAT POMPEU FABRA" and "iulaTerm".

TERMINUS constituye de hecho una estación de trabajo para la terminología. Es un sistema integral que incluye la cadena completa del trabajo terminográfico individual y en equipo: desde la búsqueda, constitución y exploración de corpus textuales hasta la gestión de glosarios y proyectos, la creación y mantenimiento de bases de datos y la edición de diccionarios en diversos formatos de impresión o electrónicos.

Organizado en varios niveles de gestión y administración, TERMINUS permite crear y gestionar equipos de trabajo articulados alrededor de un proyecto de investigación (o diversos) que puede incluir uno o varios glosarios. Permite también modelar las categorías de datos y las opciones posibles para cada proyecto, de modo que constituye un sistema flexible y adaptable a varios perfiles de usuarios y a sus necesidades concretas para la elaboración de un glosario.

El motivo fundamental que llevó al grupo a diseñar y confeccionar este sistema fue la necesidad de poder desarrollar un glosario siguiendo coherentemente la cadena de trabajo terminográfico que planteamos en los estudios teóricos: partir de corpus textuales y usar herramientas de procesamiento del lenguaje natural para llegar a detectar posibles candidatos a término para confeccionar un glosario terminológico. El horizonte era explotar al máximo la tecnología.

En su versión actual, TERMINUS consta de diversos módulos articulados:



- **Proyectos.** Permite crear y gestionar un proyecto terminológico o varios. Todo trabajo realizado en TERMINUS tiene que estar enmarcado en un proyecto determinado.
- **Fuentes.** Es el módulo de gestión de las fuentes de los términos incluidas en un proyecto terminológico, de modo que desde esta funcionalidad se pueden modificar, completar o eliminar sin necesidad de hacer cambios manuales e individuales en las fichas terminológicas de la base de datos.
- **Estructuración conceptual.** Este módulo constituye una herramienta autónoma integrada en TERMINUS que ofrece la posibilidad de crear un árbol de campo en formato gráfico, basado en nodos y relaciones, que permite estructurar conceptualmente los términos de un glosario. La interfaz gráfica facilita la construcción del árbol, que resulta muy dinámica, y su visualización.
- **Documentos.** Permite almacenar en el sistema archivos de texto que después constituirán el corpus de trabajo. Estos documentos pueden ser seleccionados previamente por el usuario e introducidos en el sistema uno a uno o por grupos, en archivos comprimidos, o se pueden extraer directamente de Internet, ya que TERMINUS incluye un buscador que interroga directamente la red sobre un determinado tema, a partir de palabras clave y/o dominios que el usuario escoge.
- **Corpus.** Permite agrupar los documentos seleccionados en el módulo anterior en uno o varios corpus textuales. Es imprescindible la creación de corpus para poder explotarlos en las fases posteriores del trabajo terminológico.
- **Análisis.** TERMINUS incorpora un paquete de análisis estadístico que permite explorar corpus con fines terminológicos mediante frecuencias, concordancias, n-gramas y cálculo de asociación entre formas. Debemos remarcar que, por el momento, no incluye un extractor automático de terminología, sino que cuenta con la intervención del usuario para la detección de términos. De hecho, la observación de las unidades de los textos por su frecuencia permite tener pistas sobre posibles términos: por ejemplo, las listas de n-gramas ordenadas por frecuencia facilitan la detección de términos poliléxicos y de posibles candidatos a términos simples, el análisis de estas unidades frecuentes en concordancia es un buen punto de observación y varias pruebas estadísticas (por ejemplo, el cálculo de asociación incluido en TERMINUS, que puntúa de mayor a menor las combinaciones de dos formas que tienden a aparecer juntas más frecuentemente de lo que debería darse por simple azar) dan también la posibilidad de detectar unidades terminológicas. En este contexto es importante señalar que, tal y como está concebido, TERMINUS es independiente de la lengua, porque en todos los casos trabaja con formas, no con lemas, de manera que no es necesario el etiquetado morfológico y sintáctico. Esta característica, además, le confiere al sistema mayor agilidad y rapidez en los diversos procesos que lleva a cabo.
- **Glosarios.** Permite declarar y gestionar los glosarios que forman parte de cada proyecto. Todos los glosarios deben asociarse a una lengua y a un área temática principales y, al crearlos, se genera automáticamente una base de datos terminológicos que se alimentará con las herramientas de otros módulos asociados.

- *Términos.* Este módulo incluye tres funcionalidades diferenciadas. En primer lugar, permite introducir en los glosarios los datos terminológicos (de forma manual o automáticamente desde el módulo de análisis de corpus), consultarlos y modificarlos. Cada ficha terminológica lleva asociadas varias categorías de datos, que el usuario puede seleccionar en función de las finalidades del glosario: lema, categoría gramatical, fuente del término, tipo de fuente, estatus del término, lengua, marca geográfica, contextos, definiciones, equivalentes, remisiones (siglas, sinónimos, etc.), colocaciones y notas lingüísticas. El sistema permite, además, utilizar automáticamente la información seleccionada del corpus textual creado previamente para incorporarla en los campos correspondientes de la ficha terminológica: lemas, fuentes y contextos. En segundo lugar, TERMINUS permite hacer búsquedas de información simples y complejas (a partir de la combinación de varios criterios) y seleccionar varios formatos de visualización de los datos. Y, finalmente, el sistema ofrece la posibilidad de exportar los datos terminológicos en diversos formatos: permite editar los glosarios en PDF, en texto plano (lo cual proporciona al usuario una mayor flexibilidad para el manejo de los datos y su presentación) o en XML, de forma que también se puede adaptar para la exportación de datos a otros gestores de terminología y a programas similares.

Además de los diferentes módulos articulados e integrados en una misma interfaz de trabajo, TERMINUS está organizado en varios niveles de administración, desde los cuales el usuario puede modelar las categorías de datos, decidir cuáles quiere incluir en función de los objetivos concretos y los destinatarios finales de cada glosario, y declarar las opciones posibles para cada categoría cerrada (por ejemplo, las categorías gramaticales o los tipos de remisiones). También desde la administración se crean y se gestionan los equipos de trabajo.

Existen tres tipos de licencias para el acceso a TERMINUS, que intentan adaptarse a tres perfiles diferentes de usuarios o grupos de usuarios: la licencia académica está especialmente diseñada para instituciones universitarias que utilizan el sistema con finalidades docentes y de investigación, ya que está concebida básicamente para el trabajo en las aulas con estudiantes; la licencia profesional está pensada para la gestión de la terminología en un organismo o en una entidad pública o privada que utiliza el sistema con finalidades profesionales; y, finalmente, la licencia individual está diseñada para el trabajo terminológico de un investigador o un equipo de investigadores autónomos.

TERMINUS es un proyecto que se va desarrollando por fases, en cada una de las cuales se intenta mejorar el producto anterior y automatizar el máximo de operaciones posibles. En la próxima fase, que prevemos iniciar en breve, intentaremos incorporar en cada módulo nuevas herramientas, por ejemplo, un extractor de terminología y un paquete de herramientas para el etiquetado de corpus textuales, de forma que se pueda ofrecer la posibilidad de trabajar con corpus sin etiquetas o con corpus etiquetados.

Hasta el momento, TERMINUS ha sido evaluado y mejorado en interacción con usuarios reales. Ha sido experimentado por los estudiantes de la Facultad de Traducción e Interpretación de la Universidad Pompeu Fabra y actualmente se están distribuyendo varias licencias académicas en universidades del Estado español y de Sudamérica.

Como vemos, pues, TERMINUS ha intentado integrar en un único sistema las herramientas y recursos informáticos útiles para mejorar y agilizar el trabajo terminográfico, sobre todo en cuanto al grado de automatización de cada proceso, básicamente en tres líneas:

- La eficiencia del proceso de trabajo, asistiendo en la tarea de documentación, consulta y adquisición de información textual.
- La identificación semiautomática de términos, incorporando técnicas estadísticas para la exploración de corpus con fines terminológicos que permiten la detección automática de unidades monoléxicas y poliléxicas candidatas a término que el usuario debe validar.
- La compleción de bases de datos terminológicos, facilitando datos que anteriormente había que reconstruir manualmente, sobre todo los datos ilustrativos de los términos.