



Centre de Llenguatges i Computació, SL



## Codificación Automática

### Presentación:

Con la extensión e implantación de las tecnologías TIC en el sector de salud y las posibilidades que estas tecnologías aportan, los diferentes sistemas de codificación están tomando cada vez una mayor relevancia e interés.

En un entorno en que los recursos son siempre insuficientes, mejorar los sistemas de gestión es importante. Un obstáculo importante reside en sus elevados costes.

La correcta codificación de los informes médicos es una de las herramientas que, con independencia de su obligatoriedad, suministra a los órganos de gestión y administración de la sanidad una información esencial para la planificación y definición de sus estrategias.

En este contexto, [Thera, Centre de Llenguatges i Computació, SL](#) presenta un sistema innovador, desarrollado con tecnologías de última generación, que permite simplificar y automatizar estas tareas. Esta innovación supone un importante avance en la aplicación de las tecnologías TIC en el sector de salud.

La complejidad de la gestión de la salud exige un sistema que facilite su control. Con esta aplicación, Thera aporta una herramienta esencial y de gran rendimiento a uno de los pilares de estas tecnologías: mejorar el conocimiento que un organismo dispone sobre una actividad.

Disponer de una información correcta y actualizada es esencial para la correcta gestión, coordinación y dirección de su actividad.

Membre de:



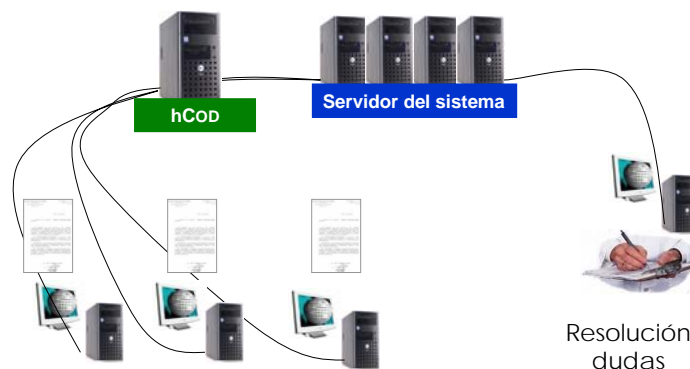
En este entorno, los diferentes sistemas de clasificación presentan una dificultad y complejidad creciente, y una mayor amplitud. Se advierte una progresiva implantación, la cual se ha convertido en una exigencia de los responsables de los sistemas de salud. Conscientes de esta necesidad, la aplicación desarrollada por Thera, hCOD, está concebida para evolucionar en el sentido de adaptarse a nuevos y más complejos sistemas de codificación, y en consecuencia, resolver las necesidades actuales y futuras que las organizaciones del sector de la salud.

### ¿Qué es hCOD?

**hCOD** es una aplicación informática desarrollada para resolver las necesidades de codificación de los informes médicos. Esta actividad quedaría automatizada en la medida en que sea obligatorio o de interés para los diferentes órganos de gestión y administración de un Centro.

La versión actual de la aplicación se adapta específicamente a la codificación de los sistemas CIE-9 MC (ICD-9 CM) y CIE-10 (ICD-10). No obstante, la aplicación puede adaptarse a extensiones particulares de éstos y a otros sistemas de codificación basados en taxonomías, y que son utilizados en diferentes sectores y especialidades del entorno médico, clínico y de asistencia sanitaria.

El motor de codificación, fundamentado en tecnologías de Inteligencia Artificial (IA) y de Procesamiento del Lenguaje Natural (PLN), analiza el contenido de los campos que deben ser codificados en los informes médicos y establece el correspondiente código sin necesidad de ninguna otra acción por parte del usuario.



El sistema ha sido desarrollado para los sistemas de codificación CIE-9 CM (ICD-9 CM) y CIE-10, (ICD-10), necesarios en los sistemas de asistencia primaria, urgencias, etc. Su uso puede extenderse a otros sistemas de codificación así como a extensiones particulares desarrolladas en un centro o sector.

**hCOD** ha sido desarrollado con el objetivo de simplificar y facilitar la actividad de codificación en los centros asistenciales, y dar soporte a los departamentos de documentación médica en sus tareas de gestión y documentación.

Para que el sistema de codificación sea de utilidad para un organismo, debe ser capaz de garantizar unos resultados con una certidumbre y confianza suficientes. Esto en el entorno de la atención médica es imposible, ya que los sistemas de codificación no son capaces de representar todas las posibilidades que en la práctica se presentan.

Para resolver esta situación, la aplicación presenta como respuesta a cada consulta no solo el código que le corresponde sino una valoración de confianza en esta decisión. Este Índice de Confianza (**Ic**) permite discriminar aquellas situaciones anómalas, indefinidas, semánticamente ambiguas o simplemente equivocadas que en la práctica se presentan. En consecuencia, permite discriminar aquellas situaciones especiales que requieren un tratamiento diferenciado.

Así pues, **hCOD** permite establecer dos vías de tratamiento de los informes: una primera en que el sistema funciona de forma automática, sin más requerimientos por parte del centro, el médico o los sistemas de gestión y documentación; y una segunda vía en que se identifican situaciones que por sus características requieren más información de la que dispone el sistema. En estas situaciones, el sistema puede actuar de diversa forma, bien solicitando más información al médico o bien identificando el informe para su posterior tratamiento específico.

De esta forma, **hCOD** permite una gestión fluida y un funcionamiento robusto y eficiente de los informes y dinamiza su gestión.

Dada la rápida evolución del conocimiento médico, no solo de la terminología aplicada sino del propio contenido de la información, los diferentes sistemas de codificación precisan de una constante y periódica revisión que incorpore al sistema aquellas innovaciones, descubrimientos o simplemente áreas de conocimiento nuevas a la codificación.

Estas continuas modificaciones exigen un sistema flexible, adaptativo y que sea capaz de asumir de forma eficiente esta inestabilidad, minimizando los costes de adaptar el sistema a estas modificaciones.

**hCOD** ha sido diseñado para resolver esta situación de inestabilidad e interinidad permanente de una forma sencilla y rápida; dispone de los recursos necesarios para simplificar al máximo estos cambios sistemáticos y periódicos de los sistemas de codificación.

A las diferentes tecnologías utilizadas en el desarrollo de esta aplicación, en las que **Thera, Centre de Llenguatges i Computació, SL** es referente y líder en España, **hCOD** incorpora una larga experiencia en el desarrollo de aplicaciones en el entorno de las tecnologías TIC.

Esta solución ha sido diseñada para adaptarse de forma fácil y rápida a las diferentes soluciones y sistemas existentes en el entorno hospitalario y asistencial para minimizar los costes de implantación y de adaptación.

En consecuencia, es una aplicación que añade a sus prestaciones un valor de estabilidad, confianza y fiabilidad esencial para su implantación dentro del sistema de salud, permitiendo una mejora sustancial de los resultados de la codificación, incrementando la confianza en éstos sistemas y su extensión a sistemas más complejos de codificación, los cuales son necesarios y están en proceso de expansión.

En consecuencia, además de ser una herramienta de codificación robusta y fiable, **hCOD** es un instrumento que permite incrementar el volumen de información y conocimiento de la actividad del centro y, en consecuencia, permitir mejoras en su gestión y administración.



Todo ello es posible gracias a 3 elementos básicos que dan a **hCOD** la potencia y capacidad suficiente como para que pueda ser considerada una aplicación de nueva generación:

- Tecnología PLN e Inteligencia Artificial
- Independencia de la lengua: *cross language technology*
- Interoperabilidad semántica

El resultado de combinar estas tecnologías es una nueva concepción de lo que debe ser una solución integral a un problema concreto y complejo al que las tecnologías tradicionales apenas pueden ayudar a resolver.

Para alcanzar este objetivo, **hCOD** dispone de los siguientes módulos:

- Identificación del idioma en que está redactado el texto.
- Análisis morfosintáctico y lematización del texto.
- Corrección ortográfica automática.
- Identificador de terminología: permite reconocer aquellos términos con contenido semántico específico dentro del área en que se lleva a cabo la búsqueda.
- Codificación: módulo que atribuye un código a un texto o fragmento de texto.
- Módulo de evaluación: establece un índice de confianza en la codificación realizada.
- Módulo de entreno: **hCOD** es un sistema de codificación que aprende por ejemplos, es decir, mediante el análisis de los historiales.

Índice de Confianza ( $I_c$ ) es un valor esencial en la aplicación práctica de hCOD, ya que es el indicador que permite definir la corrección y confianza del sistema en la codificación definida. Para índices de confianza superiores al 80%, es innecesaria una revisión posterior de la codificación establecida. Si ese resultado se considera adecuado, basta una cata periódica para verificar el correcto funcionamiento del sistema.

En líneas generales, el Índice de Confianza permite identificar situaciones de codificación que no son estándares, bien por el vocabulario utilizado en el diagnóstico bien porque el entreno del sistema no ha sido capaz, por diferentes causas, de identificar correctamente los contenidos semánticos asociados a un determinado código.

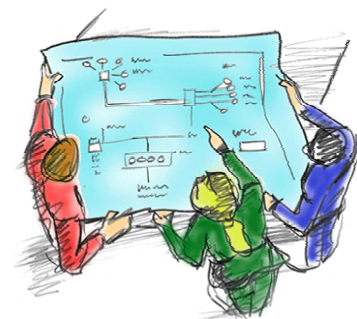
Esto permite definir dos situaciones:

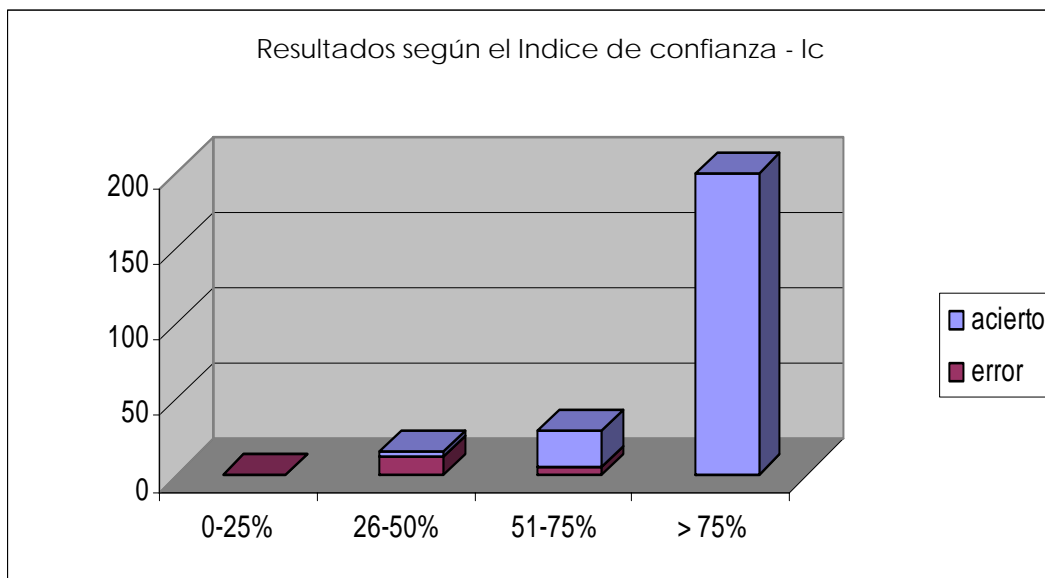
- Por una parte, la aplicación del sistema permite mejorar de forma sustancial el entreno. En consecuencia, si bien las situaciones de no codificación no disminuirán a partir de un determinado límite, si permitirán un autoajuste del índice de confianza.
- En segundo lugar, la utilización de una terminología adecuada, estandarizada y controlada en los campos de diagnóstico asegura una codificación correcta y, en consecuencia, un funcionamiento sin necesidad de revisión. Esto favorece que los usuarios se adapten al sistema, mejorando su resolución y eficiencia.

No es posible definir un código para todos los casos reales. Existen siempre situaciones ambiguas y de carencia de información en que no es posible establecer un diagnóstico.

Ello exige una aplicación flexible que sea capaz de identificar estas situaciones y no generar códigos erróneos. Estas situaciones requieren un tratamiento especial, un análisis inteligente y no deben ser tratadas como situaciones estándar.

**hCOD** soporta estas situaciones y permite la existencia de un determinado nivel de ambigüedad.





La incorporación de un sistema de Codificación Automática en un entorno definido, presenta una interacción positiva con el centro, generando una dinámica que hace tender a la uniformización de la terminología utilizada por el centro.

En efecto, los sistemas basados en el aprendizaje automático aportan una ventaja de eficiencia a los usuarios que utilizan una misma terminología y un *vocabulario controlado*, al dar a este conjunto una mejor respuesta, esto reduce su esfuerzo en la codificación de los informes.

Los éxitos, considerando como tales las codificaciones automáticas con un índice de confianza suficiente, suponen la automatización del sistema mientras que las divergencias implican un sobrecoste de confirmación y verificación.

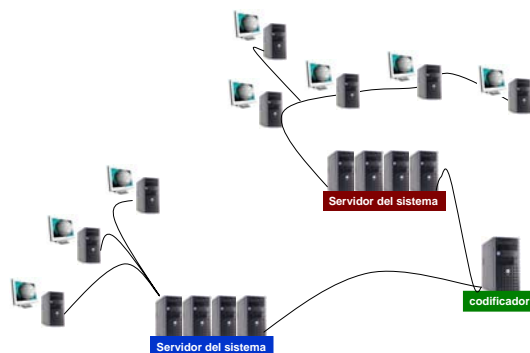
En esta situación, la tendencia es la de buscar aquellas expresiones que ya son comprendidas por el sistema, y una adecuación a éste.

A medio plazo, esto supone una mejora en el rendimiento del sistema y una mejor resolución en la codificación. La responsabilidad de la codificación se trasfiere al sistema informático con lo que el médico puede dedicar su esfuerzo a la realización de su trabajo, cediendo los aspectos administrativos y de gestión al sistema informático.

## CARACTERÍSTICAS TÉCNICAS

**hCOD** está desarrollado bajo la plataforma Linux y trabaja independientemente del sistema de gestión. Desarrollado con tecnología WEB, el sistema puede adaptarse fácilmente a cualquier otro sistema y plataforma. Recibe la información a tratar por mediación de la red (Intranet o Internet), procesándola sin interferir con los demás servidores o usuarios del sistema.

Está diseñado para realizar sus funciones de forma automática y dispone de las interfaces necesarias para su entreno, activación y control por parte del responsable del sistema. Su velocidad y prestaciones son elevadas, y ha sido diseñado y realizado para ser escalable en función del volumen de informes que debe codificar.



**Thera** participa activamente en el desarrollo, extensión e implantación de los estándares de interoperabilidad médica, y colabora con la definición de un estándar de interoperabilidad semántica que permita la aplicabilidad de sus sistemas en el entorno de salud de forma sencilla, transparente y eficiente.

La aplicación desarrollada por Thera para la codificación automática se fundamenta en tecnologías de IA, PLN y *Cross Language*, lo que le permite resolver de forma directa la existencia de diferentes idiomas en la práctica médica del Estado. Por lo tanto, puede instalarse para la codificación de Informes médicos con independencia de la lengua en que estos informes estén redactados.

## ¿Cómo funciona hCOD?

Como ya se ha indicado, para el desarrollo de esta aplicación se han utilizado tecnologías de Procesamiento del Lenguaje Natural y de Inteligencia Artificial, esto le permite dar una respuesta correcta en el mayor número de situaciones, ser flexible y adaptarse a una realidad compleja y difícil.

Para poner su puesta en marcha, el sistema analiza los precedentes (histórico), construyendo para cada uno de los códigos, un modelo semántico del lenguaje utilizado en su definición.

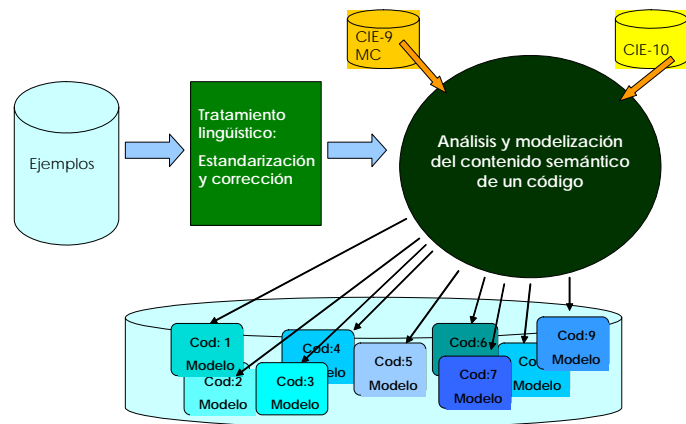
En la construcción de este modelo son esenciales dos elementos: un lenguaje estandarizado, obtenido mediante la aplicación de tecnologías de PLN (lematización, corrección ortográfica, etc.) y un conocimiento del léxico del sector médico.

Una vez establecido el correspondiente modelo de uso del lenguaje para cada uno de los códigos del sistema de codificación, **hCOD** puede entrar en producción.

Cuando se le suministra un diagnóstico, en primer lugar estandariza su contenido y luego compara éste con los diferentes modelos existentes.

De esta forma el sistema establece la proximidad entre el texto introducido y los diferentes modelos, y después establece a cual de ellos pertenece la formulación introducida.

Funcionamiento del **Motor de codificación**:

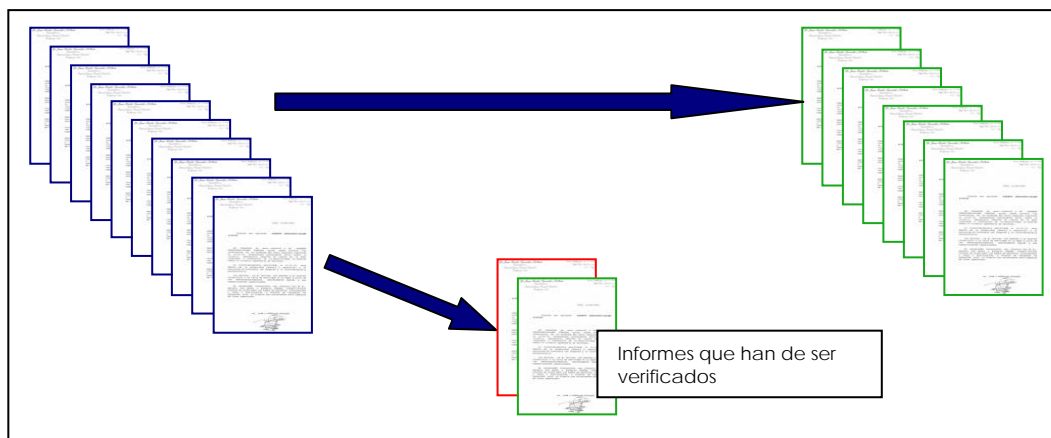


El sistema no da una única respuesta, está preparado para aportar una relación de posibles códigos en base a este análisis. Cada una de las soluciones propuestas tiene un grado de proximidad que es utilizado para construir el Índice de Confianza. Cuando hablamos de un acierto del 91% queremos indicar que el código a que el sistema atribuye mayores posibilidades es el correcto.

Dado que las posibilidades (confianza) son variables, cuando este Índice no alcanza el nivel de suficiencia (75%), lo que ocurre en un 20% de los casos, el sistema permite codificar sin necesidad de revisión un importante volumen de informes: un seguimiento de la estabilidad del sistema mediante catas es suficiente para garantizar la corrección de las codificaciones realizadas.

De esta forma, se reduce en un 80% el volumen de codificaciones que deben ser realizadas o supervisadas por parte del Centro, sus documentalistas médicos, o por los propios médicos, en el proceso de diagnóstico.

Como ya se ha indicado, el sistema aporta otras soluciones, que pueden ser propuestas en las situaciones en que el Índice de Confianza no adquiere un valor suficiente, simplificando la tarea de codificar por parte del personal médico del Centro.



[Thera, Centre de Llenguatges i Computació, SL](#) está especializada en el desarrollo de aplicaciones basadas en el Procesamiento del Lenguaje Natural, el tratamiento semántico de la información y el desarrollo de aplicaciones tecnológicas de alta calidad.

La actividad de Investigación desarrollada por [Thera](#) le permite estar en vanguardia de la tecnología, aportando soluciones a las empresas y organizaciones que precisan tratar grandes volúmenes de información y organizarla, clasificarla o localizarla.

Dentro del sector de Salud, [Thera](#) aporta su experiencia, investigaciones y sus conocimientos en el Procesamiento del Lenguaje Natural (PLN), para desarrollar aplicaciones centradas en la [Interoperabilidad Semántica](#), los diferentes sistemas de codificación, y el desarrollo y aplicación de sistemas de [representación del conocimiento y ontologías](#).

Con esta aplicación, [Thera](#) da un paso en el desarrollo de sistemas de codificación inteligentes, poniendo a disposición de los usuarios una herramienta que permite, además del cumplimiento de las necesidades administrativas de codificación, la posibilidad de explotar el conocimiento disponible en sus Bases de Datos de Informes. Gracias a esta aplicación, se pueden obtener estadísticas e informes de la actividad médica del centro y la explotación de esta información en la toma de decisiones por parte de los responsables de gestión y dirección.

[Thera](#) pertenece a la [Xarxa d'Innovació Tecnològica](#) (XIT) de la Generalitat de Catalunya y centra su esfuerzo en el desarrollo de aplicaciones innovadoras, incorporando los descubrimientos, aportaciones y novedades desarrolladas en su departamento de investigación, transfiriendo tecnología de los Centros de Investigación y aportando este conocimiento a aplicaciones que permitan mejorar la competitividad de nuestros clientes.

Desde el año 2001, [Thera](#) desarrolla su actividad de transferir el conocimiento y la tecnología desarrollada en las universidades y en su propio Departamento de Investigación, transformando este conocimiento en herramientas, aplicaciones y productos de un elevado valor añadido, innovadores y que respondan a las necesidades del mercado.

De esta forma, [Thera](#) participa de forma activa en la construcción de la Sociedad de la Información y en el desarrollo de las Tecnologías de la Información y la Comunicación (TIC), colaborando con diferentes Instituciones, Centros de Investigación, Universidades y otras Empresas o Centros Tecnológicos en el desarrollo y consolidación de los canales de acceso al mercado de estas tecnologías.

---

Membre de:



## TECNOLOGÍAS LINGÜÍSTICAS

**Thera**, empresa líder en el mercado nacional y europeo en el desarrollo de las tecnologías de la lengua, dispone de una amplia experiencia en el desarrollo de recursos básicos y aplicaciones de procesamiento básico del lenguaje:

*Lematización, análisis morfosintáctico, detección de terminología y Entidades Citadas, construcción de ontologías y sistemas de representación del conocimiento, y 'pippe linnes' de tratamiento lingüístico* se ubican como elementos básicos que permiten un análisis y procesamiento del texto de forma rápida y eficiente.

Estos procesos de análisis permiten la explotación del conocimiento aportado a la información en este proceso para una resolución y exactitud que no tiene competencia en un mercado basado en tecnologías convencionales.

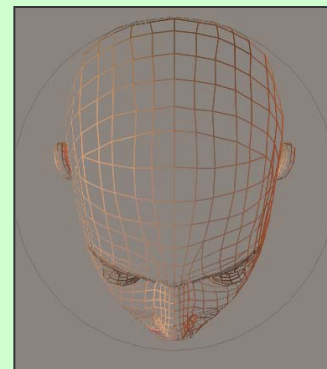
**Thera** ha desarrollado tecnología propia para el desarrollo de sus aplicaciones comerciales y mantiene un nivel de investigación y soporte, que permite evolucionar sus productos incorporando en cada momento las tecnologías y recursos más avanzados. Esta práctica convierte a **Thera** en una empresa de primera línea a nivel internacional, y a sus soluciones en la mejor opción de un mercado en que la I+D son esenciales.

### **HS-Morfo**<sup>®</sup> Lematizador de alta velocidad.

Thera dispone, entre sus herramientas básicas, de la primera versión de un lematizador de altas prestaciones, utilizable como herramienta básica de procesamiento en diferentes aplicaciones (Ontology<sup>®</sup>, KnowBoard, CeRes, EnCit, ...).

### **WordNet y EuroWordNet**

En el momento actual, en que se inicia el desarrollo de ontologías como sistemas de representación del conocimiento, la red semántica *WordNet* y su versión multilingüe *EuroWordNet* son un elemento esencial en el desarrollo de aplicaciones inteligentes.



Membre de:



## El equipo humano

La tecnología desarrollada por Thera es fruto de la colaboración de un excepcional equipo humano que une a su dinamismo, juventud y entusiasmo, una voluntad de cooperación, una experiencia multidisciplinar, una visión abierta y de futuro, capaz de superar las dificultades y desafíos de una nueva tecnología.

Ingenieros, lingüistas y matemáticos se unen de forma creativa en el diseño y desarrollo de aplicaciones innovadoras, recursos y herramientas que suman los conocimientos y especialización de cada tecnología, creando productos únicos y excepcionales tanto en su calidad como en sus prestaciones.



Membre de:



### Principales líneas de productos y servicios:

- Extracción de información
- Clasificación automática
- Etiquetado semántico
- Q/A (búsqueda de respuestas)
- Identificación y clasificación entidades
- Generación automática de informes
- Indexación y recuperación de información
- Codificación automática
- Generación de ontologías
- Corrección ortográfica



**Thera**, Centre de Llenguatges i Computació, S.L.  
Parc Científic - Universitat de Barcelona  
c/ Adolf Florensa s/n Ed. Florensa  
08028-BARCELONA

tel.: 93 403 45 58

fax.: 93 403 46 78

<http://www.thera-clic.com>

e-mail: [thera@thera-clic.com](mailto:thera@thera-clic.com)