

TRADUCCIÓN AUTOMÁTICA ENTRE LENGUAS EUROPEAS

CONTACT DETAILS:

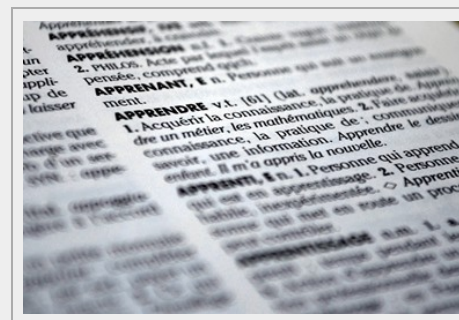
OTRI - Área de Relaciones con la Empresa
 Universidad de Alicante
 Tel.: +34 96 590 99 59
 Email: areaempresas@ua.es
<http://innoua.ua.es>

ABSTRACT

El grupo Transducens del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante es experto en el desarrollo de software para traducción automática entre pares de lenguas emparentadas, con énfasis en lenguas románicas y aplicable a pares de lenguas eslavas y escandinavas de Europa, tanto mayoritarias como minoritarias.

Las principales características son: buena combinación entre velocidad y calidad de traducción, posibilidad de utilización en web y capacidad para desarrollar software en código abierto o cerrado.

El grupo busca aplicar su know-how en proyectos específicos de tratamiento de lenguaje natural.



TECHNICAL DESCRIPTION

El grupo Transducens del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante es experto en el desarrollo de software para el tratamiento del lenguaje humano, en particular para la traducción automática y pronunciación de lenguas emparentadas, con énfasis en lenguas románicas de Europa.

El grupo está muy interesado en la preservación de culturas locales y, por tanto, presta especial atención a las lenguas de uso minoritario en Europa. El grupo tiene el know-how y la capacidad para desarrollar motores de traducción automática con las siguientes características:

- **Aplicables a la traducción automática entre pares de lenguas emparentadas**, como por ejemplo entre lenguas románicas (español, francés, italiano, catalán...), lenguas eslavas (polaco, checo, eslovaco; o búlgaro, macedonio) y lenguas escandinavas (danés, sueco, noruego, islandés).
- **Modulares**: el motor está constituido por módulos independientes que se encargan de cada tarea, la mayoría de los cuales se basan en técnicas de estados finitos.
- **Traducción automática muy veloz**, gracias a la utilización de técnicas de estados finitos. Traducción aproximada de unas 10.000 palabras por segundo (con un ordenador de sobremesa común).
- **Altos niveles de calidad de traducción**, entendiendo como tal el porcentaje de texto traducido automáticamente que no tendrá que ser corregido porque se trata de una traducción razonablemente correcta y aceptable del original. Los niveles de calidad oscilan entre el 85% y el 95% del texto correctamente traducido.
- **Fácil integración** de los motores desarrollados en aplicaciones ejecutadas en Internet, diferentes a la traducción automática.
- **Navegación traducida**: se modifican automáticamente todos los enlaces para que apunten a las traducciones de las páginas correspondientes.
- Los textos pueden estar en cualquiera de los siguientes **formatos**: ANSI (texto plano), HTML (lenguaje de marcas de hipertextos usados en las páginas de Internet) y RTF (formato de texto enriquecido de Microsoft).
- **Software abierto**: en la actualidad se está procediendo a una reimplementación completa de todos los módulos con la intención de liberarla como software abierto (licencia de tipo GPL).

- Empleo de caracteres universal (UNICODE). Actualmente se están migrando todos los códigos a UNICODE para evitar incompatibilidades en traducciones entre lenguas de diferentes regiones europeas.

PROCESO DE DISEÑO

El diseño es muy sencillo: se basa en módulos que van realizando la traducción por etapas. Se pueden distinguir siete etapas básicas:

- a. Separación del texto de la información de formato del documento.
- b. Análisis morfológico de las palabras y las locuciones del texto.
- c. Elección (fundamentada en el contexto y mediante procedimientos estadísticos) de un único análisis morfológico en el caso de las palabras que tengan varios significados.
- d. Determinación de determinadas estructuras sintácticas sencillas de más de una palabra que exijan un trato especial (concordancia de género y número, cambios de preposición, reordenaciones), producción de la estructura correspondiente en la lengua de llegada y consulta del diccionario bilingüe de equivalencias.
- e. Generación de las formas conjugadas y flexionadas de las palabras, así como de las locuciones en la lengua de llegada.
- f. Apostrofación, contracciones y guionaje de palabras.
- g. Restitución de la información de formato para conseguir un documento traducido con un formato tan parecido como sea posible al del original.

Este funcionamiento es razonablemente potente pero sencillo, lo que permite una programación muy eficiente de las etapas y, como resultado, una velocidad de traducción de decenas de miles de palabras por segundo.

APLICACIONES DESARROLLADAS

Se han desarrollado varias aplicaciones comerciales que utilizan este know-how. Por ejemplo:

- Traductor InterNOSTRUM castellano-catalán (<http://www.internostrum.com>).
- Traductor Universia español-portugués (<http://traductor.universia.net>).
- Sistema de ayuda para la pronunciación en voz alta del idioma valenciano (<http://sao.dlsi.ua.es/>).

TECHNOLOGY ADVANTAGES AND INNOVATIVE ASPECTS

ASPECTOS INNOVADORES

Aunque existen numerosos desarrolladores de productos de traducción automática, son pocos los que tienen como éste la capacidad de trabajar con lenguas minorizadas.

VENTAJAS

- Una buena combinación entre velocidad y calidad de traducción.
- El empleo de modelos de estados finitos confiere una gran velocidad a los sistemas desarrollados.
- El hecho de tratarse de un software en código abierto hace posible una mejor adaptación a las necesidades concretas de cada aplicación, además de permitir una mejor depuración de errores.
- La posibilidad de aplicación de los motores desarrollados en web.

CURRENT STATE OF DEVELOPMENT

El know-how está disponible para ser transferido y utilizado para desarrollar nuevas aplicaciones. Se han desarrollado varias aplicaciones comerciales que utilizan este knowhow.

MARKET APPLICATIONS

El know-how se puede aplicar al desarrollo de:

- Motores de traducción automática entre pares de lenguas emparentadas, como por ejemplo entre lenguas románicas (español, francés, italiano, catalán...), lenguas eslavas (polaco, checo, eslovaco; o búlgaro, macedonio), lenguas escandinavas (danés, sueco, noruego, islandés).
- Motores de traducción automática entre pares de lenguas minoritarias europea, como por ejemplo occitano, bretón, etc.

- Sistemas de ayuda a la lectura en voz alta de textos en lenguas con sistemas de escritura aproximadamente fonética.
- Motores utilizables en aplicaciones web (traductores on-line, navegación por internet con traducción) y para ser usados tanto en código abierto como cerrado.

COLLABORATION SOUGHT

- Tipo de cliente buscado: industrias, universidades y centros tecnológicos.
 - Sectores: Tratamiento del lenguaje humano, software de traducción, aplicaciones web.
- El grupo de investigación está interesado en la aplicación de su know-how en proyectos específicos.

INTELLECTUAL PROPERTY RIGHTS

Todos los códigos, diccionarios, motores, etc. utilizados han sido desarrollados por el grupo de investigación y están protegidos por know-how.

MARKET APPLICATION (1)

Informática, Lenguaje y Comunicación