

Tecnologías de la Información y de la Comunicación en euskera (TIC's)



Araceli Diaz de
Lezana

Responsable de Planificación Lingüística, Viceconsejería de Política Lingüística

Hace algunos años, escuchamos por primera vez el concepto "sociedad de la información", y, al mismo tiempo, empezamos a hablar de las "nuevas tecnologías". Internet, es decir, la Red de redes, nos ha cambiado la vida: el modo de trabajar, la manera de hacer negocios, el modo de relacionarnos con los demás, la manera de recibir servicios, el modo de recibir la información... Junto con Internet, surgió un concepto nuevo, la "realidad virtual", un espacio que llega hasta nosotros a través del ordenador.

Gracias a este nuevo mundo, han surgido nuevas necesidades, impulsadas en gran medida por Internet, y, necesariamente, el desarrollo tecnológico se ha centrado particularmente en un ámbito: el desarrollo de las tecnologías lingüísticas. Hoy en día, este tipo de tecnologías se denominan "Tecnologías de la Información y de la Comunicación (TIC)". A nivel mundial, estas tecnologías se conocen como *Information Technology* (IT). Veamos cómo define este concepto la *Information Technology Association of America*: "the study, design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware...". *Information Technology is a general term that describes any technology that helps to produce, manipulate, store, communicate, and/or disseminate information.* [Fuente: Wikipedia] (Estudio, diseño, desarrollo, implementación, mantenimiento o gestión de sistemas por ordenador utilizando hardware y software... Las Tecnologías de la Información es un término general que define las tecnologías que ayudan a crear, manipular, almacenar, comunicar y/o difundir información).

Así, vivimos en la era de la información y de la comunicación, y, por ello, las universidades, los centros de investigación y las empresas están desarrollando tecnologías que realizan un tratamiento especial de las lenguas, es decir, "la infoingeniería lingüística" o las "tecnologías lingüísticas (*Language Technologies*)". Han surgido nuevas profesiones y mercados, y, gracias a ellos, dichas tecnologías están avanzando de forma considerable. Fundamentalmente son éstas:

- El reconocimiento de la voz o el habla (**Speech-To-Text** y **Automatic Speech Recogniton**)
- La síntesis de la voz o el habla (**Text-To-Speech**)
- El reconocimiento óptico de caracteres (**Optical Character Recognition**)
- La traducción automática (**Text-To-Text** y **Speech-To-Speech**)
- La compresión de datos de voz
- "Data / audio mining" (gestión de grandes cantidades de información)

En la comunicación actual, la tecnología se ha convertido en una prioridad, ya que representa el nivel de desarrollo de un pueblo. Gracias a la tecnología nos comunicamos fácilmente, y desaparecen los obstáculos geográficos y temporales. Sin embargo, la influencia que puede llegar a tener la tecnología siempre está en manos del usuario, es decir, depende del uso que éste haga de ella, de si pretende obtener información, buscar entretenimiento, trabajar, aprender, recibir formación... o, por el contrario, utilizarla con otros fines: realizar operaciones fraudulentas, controlar a personas y perjudicarlas, manipular y falsificar datos...

De todos modos, el desarrollo de las TIC está condicionado por los usuarios, porque ellos son los que deciden en todo momento qué tipo de productos y contenidos desean obtener. Por todo ello, se considera que las tecnologías tienen una gran influencia en el proceso de construcción de la sociedad. Es importante formar a la audiencia, ya que cuanto más utilicen los usuarios su creatividad, mejores serán los productos que se desarrollarán, y menor el número de exclusiones de éstos en el mercado.

Conscientes de la importancia de las Tecnologías de la Información y de la Comunicación, los gobiernos y las administraciones están impulsando políticas enfocadas a desarrollar dichas tecnologías. Por un lado, observan que como consecuencia del desarrollo de éstas se están produciendo cambios sociales, económicos y políticos, y, por otro, son conscientes de que la utilización de las tecnologías ofrece una capacidad casi ilimitada para obtener, recabar, procesar y compartir información.

1. Las Tecnologías de la Información y de la Comunicación en euskera.

En el marco de la "Iniciativa Euskadi 2000Tres", el Gobierno Vasco promovió diversas actuaciones con el objetivo de que la sociedad vasca se integrase de lleno en la Sociedad de la Información. Para ello, puso el espacio digital al servicio de la ciudadanía.

Así mismo, el Gobierno Vasco ha puesto en marcha varios planes para desarrollar las Tecnologías de la Información y de la Comunicación en Euskadi e integrar el euskera en ellas:

1999, Plan General de Promoción del Uso del Euskera

El plan estableció los siguientes objetivos para la siguiente década: promover la industria de la ingeniería del euskera; priorizar la localización de los productos de gran difusión en el mercado; crear un comunidad de usuarios preparados y cualificados; crear plataformas o redes de edición y traducción, y poner diccionarios electrónicos al alcance de todos; desarrollar sistemas multimedia para aprender euskera y crear una infraestructura de las herramientas básicas para el análisis lingüístico.

2001-2004, Plan de Ciencia, Tecnología e Innovación

Dentro del programa de las industrias lingüísticas, el objetivo del Plan era impulsar los siguientes campos prioritarios:

- Desarrollo de las técnicas necesarias para la ingeniería lingüística.
- Desarrollo de las herramientas lingüísticas.
- Promover la utilización de instrumentos avanzados y la experimentación de plataformas de capacitación.
- Recabar, recuperar y buscar información.

- Reconocimiento automático de la voz, compresión digital de la voz y conversión del texto en voz.
- Funcionalidades útiles para el tratamiento de textos.
- Tratamiento del texto en el contexto de la ofimática y en la conversión de texto a texto (traducción).

2002-2005, Plan Euskadi en la Sociedad de la Información (PESI)

En el apartado "Euskadi en Red" se mencionan las iniciativas a favor del euskera: "Desarrollo de herramientas lingüísticas en la red, desarrollo y promoción de contenidos en euskera en la red, promoción de proyectos innovadores en el campo de las nuevas tecnologías, tecnologías de traducción automática, voz y tratamiento de textos aplicados al euskera".

2003-2005, Plan de Informática y Telecomunicación

Este plan reunió los proyectos sobre los sistemas de información surgidos de las iniciativas mencionadas en los planes anteriores, además de otros proyectos, teniendo en cuenta las siguientes líneas de actuación:

- Mejora de la gestión interna
- Administración electrónica
- Infraestructuras y organización

PCTI 2010, Plan de Ciencia, Tecnología e Innovación

En los apartados 8.1.7 y 8.1.8 se establecen los objetivos de las tecnologías de la información y de las industrias lingüísticas.

2008, Plan Euskadi en la Sociedad de la Información 2010: Agenda Digital de Euskadi

En el apartado "Euskadi en Red" se especifica lo siguiente:

Asegurar la presencia de Euskadi en la red, preservando en los nuevos entornos digitales su cultura y su lengua y poniendo en el escaparate de Internet la oferta de sus recursos turísticos.

Ello exige articular una línea de actuación cuyo objetivo sea triple:

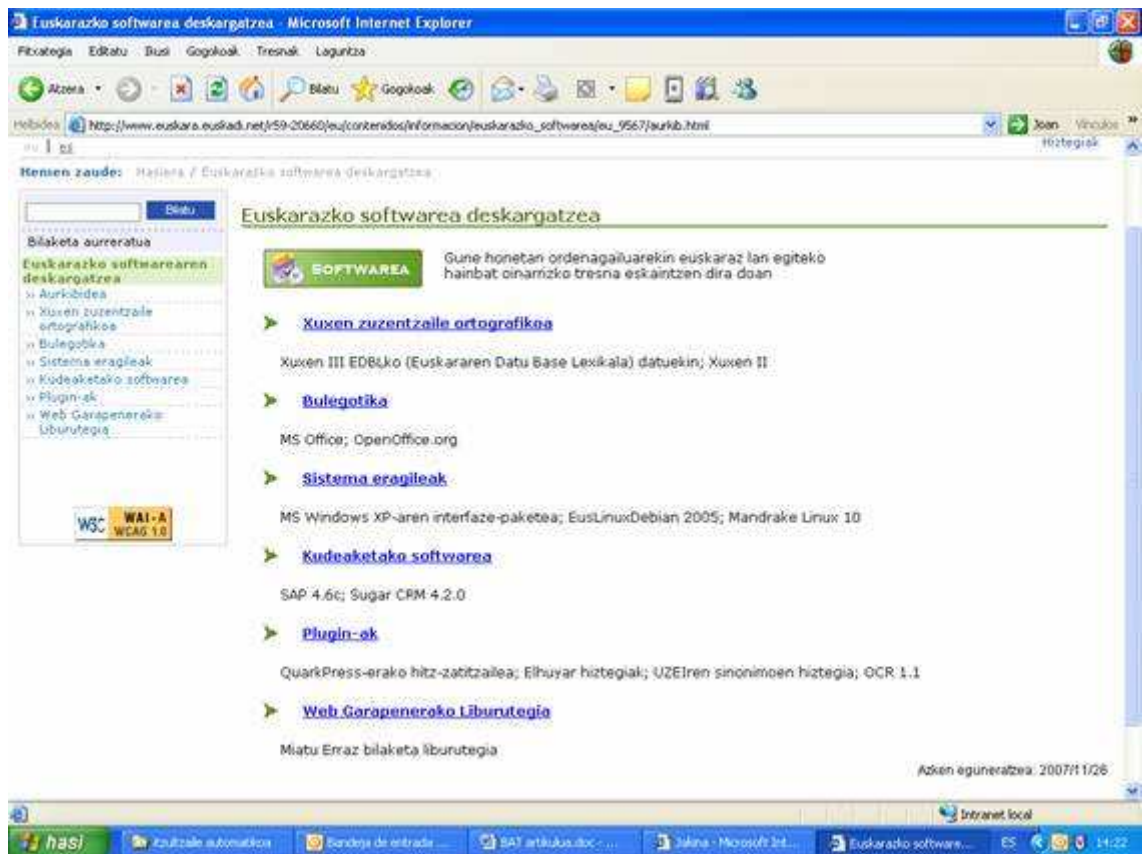
- Por un lado, mantener una presencia rica, dinámica y continua del Euskera y la cultura vasca en Internet y los medios digitales, que permita preservar y divulgar la identidad cultural vasca.
- Por otro, garantizar a la ciudadanía vasca una experiencia digital plena en Euskera.
- Finalmente, dar a conocer los atractivos y los recursos turísticos que ofrece Euskadi a sus visitantes.

Junto con estos planes, hay que destacar, por un lado, la convocatoria de subvenciones para promover las TIC que promueve todos los años la Viceconsejería de Política Lingüística del Departamento de Cultura, especialmente la destinada a financiar los productos para el ámbito de las TIC en euskera, y, por otro lado, los programas del Departamento de Industria cuyo objetivo es impulsar la investigación: ETORTEK, SAIOTEK e INTEK. Estos programas están dirigidos a los agentes de la Red Vasca de Ciencia, Tecnología e Innovación. Las Diputaciones Forales también otorgan subvenciones para promover dichas tecnologías.

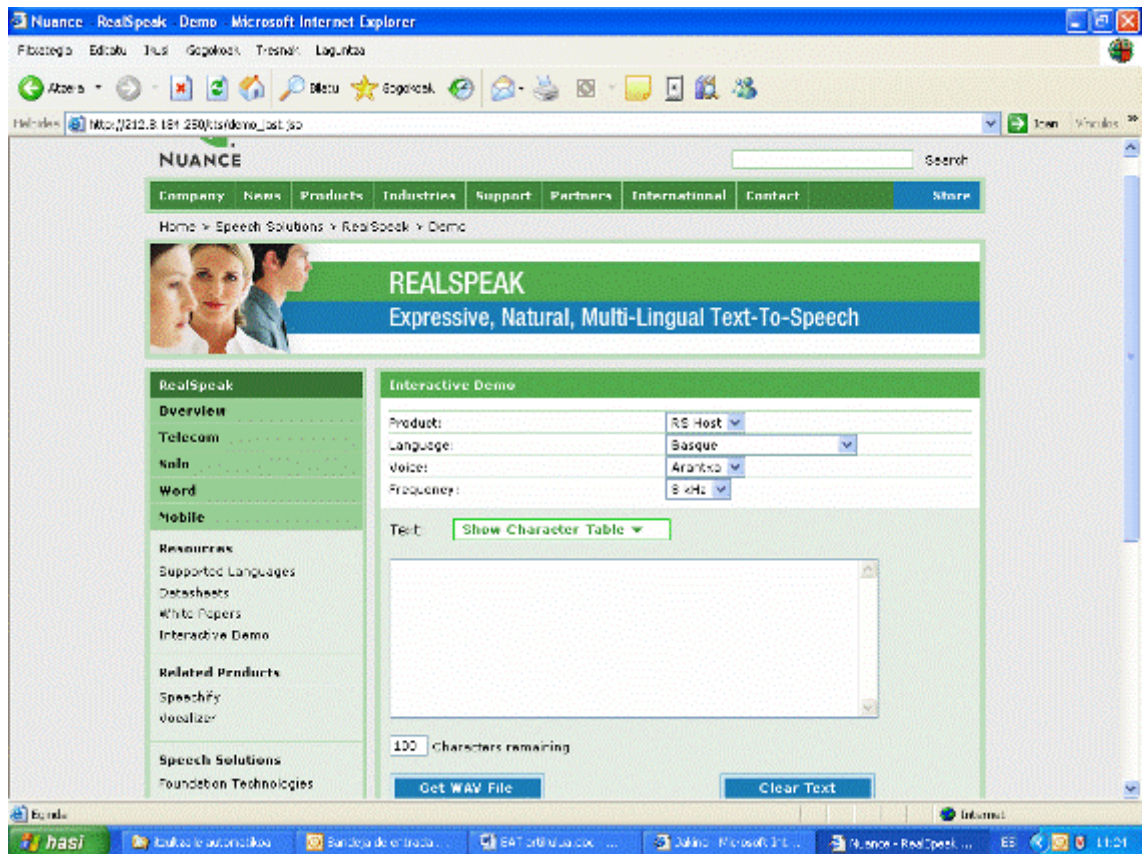
En los últimos años se han desarrollado muchos proyectos para integrar el euskera en las TIC por iniciativa de las administraciones. A ello hay que sumar el gran trabajo que han realizado y están realizando los agentes tecnológicos vascos en este ámbito.

A continuación, mencionamos de manera resumida y general los productos y recursos que se han desarrollado en euskera:

- En la dirección www.euskadi.net/euskara_soft se puede disponer de manera gratuita de varios productos: los sistemas operativos Windows Vista y XP (paquete de interfaz) y Linux; los paquetes de ofimática MS Office 2007 y 2003 (paquete de interfaz) y OpenOffice.org; plugins (diccionarios Elhuyar, diccionario de sinónimos de UZEI, el divisor de sílabas o hyphenator) para QuarkPress; sistema OCR); la biblioteca de búsqueda Miatu para el desarrollo de páginas web; todas las versiones del corrector ortográfico Xuxen (incluida la de código abierto); y software de gestión para empresas (cinco módulos de la versión SAP 4.6c; Sugar CRM 4.2.0).



- Síntesis y reconocimiento de la voz o del habla: el Departamento de Cultura del Gobierno Vasco llegó a un acuerdo con la empresa Nuance (antes Scansoft) para desarrollar los motores para el euskera www.nuance.com.



- Recursos lingüísticos: diccionarios electrónicos (Diccionario Unificado, Elhuyar, 3000, Morris, Euskal Hiztegia, Sinónimos, el diccionario enciclopédico Harluxet...); bases de datos lexicales (EDBL); corpus (corpus estadístico, corpus de ciencia y tecnología, corpus literario); recursos para integrar en las tecnologías de la voz (bases de datos acústicas, léxico fonético, corpus textuales); memorias de traducción; corpus etiquetado de los boletines oficiales; banco terminológico público (Euskalterm); toponimia (base de datos de la CAV) y onomástica (los nomenclátors de la Real Academia de la Lengua Vasca)...
- Herramientas de la lengua: buscadores, lematizadores, analizadores, corrector, etiquetadores, tesauros, herramientas para realizar extracciones automáticas, herramientas para la traducción automática...
- Formación electrónica: cursos on line (masters, cursos...).
- Foros, sitios de debate, boletines electrónicos...
- Contenidos digitales (prensa, revistas...)
- ...

Además, en julio de 2008, la Viceconsejería de Política Lingüística adjudicó mediante concurso público un contrato para desarrollar una herramienta de traducción automática de textos generales de castellano a euskera. Los trabajos se prolongarán durante quince meses a partir de la fecha de adjudicación del contrato, y la herramienta final podrá ser utilizada:

- Desde el sitio www.euskara.euskadi.net se podrán traducir textos breves y sitios web.
- Desde la intranet del Gobierno Vasco se podrán traducir documentos en diferentes formatos, que serán corregidos por profesionales y que, a su vez, servirán para mejorar el sistema.

Después de este primer paso, se prevé desarrollar nuevas direcciones de traducción: euskera-castellano, euskera-inglés y euskera-francés. Las tentativas realizadas hasta el momento en euskera se han llevado a cabo de castellano a euskera, teniendo en cuenta que es importante que la lengua de origen sea una lengua con gran desarrollo tecnológico.

2. Inventario de las Tecnologías de la Información y de la Comunicación (TIC) en euskera

La Viceconsejería de Política Lingüística (VPL) definió el proyecto para el período 2005-2009 a través de la publicación *Futuro de la política lingüística* (Gobierno Vasco, Vitoria-Gasteiz, 2005, ISBN 84-457-2407-X).



De todas las medidas que se indican en dicha publicación, la nº 23 dice lo siguiente:

Desarrollo de recursos, herramientas y aplicaciones nuevas en euskera en el campo de las nuevas tecnologías de la Información y de la Comunicación. Asimismo, incremento de la presencia del euskera en Internet, creando contenidos digitales... En este marco se considera necesario analizar y radiografiar la presencia actual del euskera en ese ámbito, así como los recursos en euskera ya disponibles en estas nuevas tecnologías, y definir el camino a seguir en el futuro.

Por todo ello, la VPL empezó a completar el **Inventario de las TIC's** en 2006. A través de esta iniciativa, la VPL quería lograr los siguientes objetivos:

- Obtener un diagnóstico y un mapa de las iniciativas realizadas y en proceso.
- Reconocer los puntos fuertes y los puntos débiles de lo realizado hasta ahora y definir las prioridades de futuro.
- Identificar los proyectos y las iniciativas que podrían concordar con el plan estratégico del Gobierno Vasco, teniendo en cuenta siempre el valor añadido y la eficacia de éstos.
- En definitiva, obtener información cualificada para incrementar la presencia del euskera en las nuevas tecnologías y dar a conocer esa información a través del Sistema Vasco de Indicadores:
 - La presencia del euskera en las nuevas tecnologías
 - La aportación de las nuevas tecnologías al euskera

El primer paso consistió en organizar el ámbito de las TIC. Así, para integrar los proyectos en el Inventario fue necesario calificarlos en un subcampo específico:

Clasificación de las Tecnologías de la Información y de la Comunicación (Inventario de las TIC)

www.euskara.euskadi.net/ikt_inbentarioa

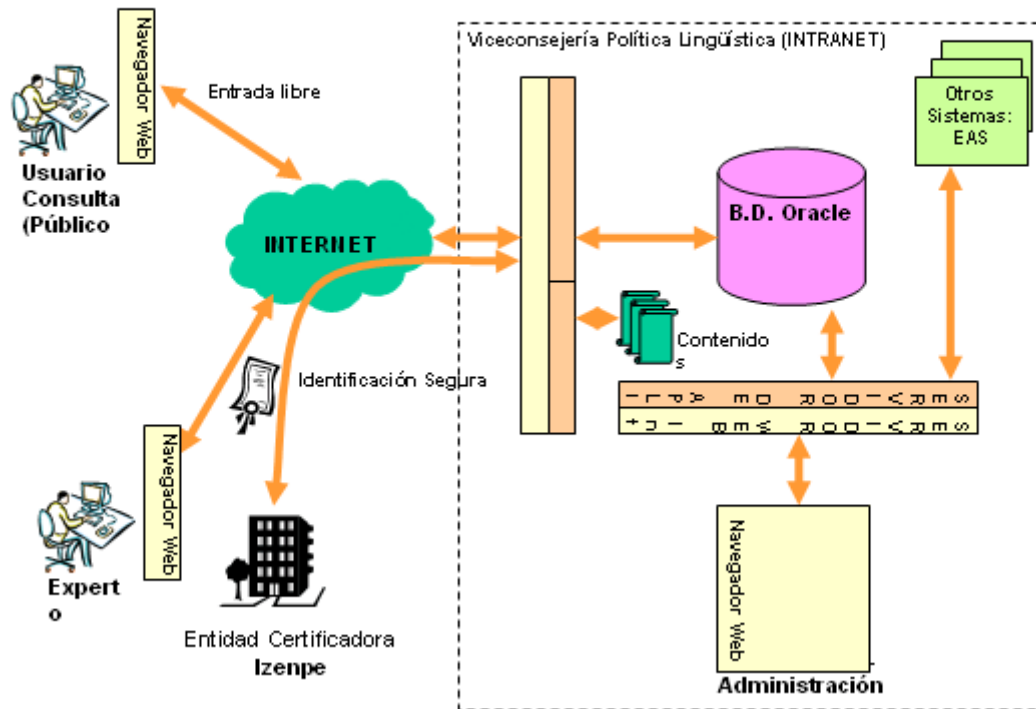
RECURSOS Y HERRAMIENTAS LINGÜÍSTICAS	GESTIÓN DE LA INFORMACIÓN Y DEL CONOCIMIENTO
<ul style="list-style-type: none"> - recursos de voz - glosarios - diccionarios - correctores - lematizadores - edición de documentos plurilingües - sistemas OCR - bancos terminológicos - ontologías - analizadores - etiquetadores - tesauros - bases de datos léxicas - corpus - otro tipo de herramientas de trabajo 	<ul style="list-style-type: none"> - recuperar documentación - extraer información - realizar resúmenes automáticos - clasificar la documentación - reconducir la documentación - almacenar información - filtrar documentación - crear textos automáticamente - otros
FORMACIÓN / E-LEARNING	SOFTWARE EN EUSKERA
HERRAMIENTAS DE AYUDA A LA TRADUCCIÓN	<ul style="list-style-type: none"> - ofimática - gestión empresarial - herramientas de utilización general - enseñanza - idioma - tiempo libre - herramientas de trabajo de Internet - información técnica relacionada con el software
<ul style="list-style-type: none"> - memorias de traducción - herramientas de traducción automática - otras 	<ul style="list-style-type: none"> - otros
SISTEMAS INTERACTIVOS AVANZADOS	INDUSTRIA DE CONTENIDOS (crear contenidos digitales en euskera)
<ul style="list-style-type: none"> - reconocimiento automático de la voz - sistemas para contestar preguntas - identificación biométrica a través de la voz - síntesis de la voz 	INCLUSIÓN ELECTRÓNICA (aplicaciones para los que tienen alguna discapacidad)
	OTROS

El siguiente paso consistió en contactar con los agentes tecnológicos de la CAV, con objeto de recabar información sobre las TIC en euskera, ya que su aportación resultaba totalmente necesaria e imprescindible. De este modo, en un principio el Inventario de las TIC en euskera se puso en marcha con la información aportada por 15 agentes: el Grupo DELi de la Universidad de Deusto, La Fundación Elhuyar, la Organización Vasca GNU, Eleka Ingeniaritza Linguistikoa S.L, el Grupo IXA de la Universidad del País Vasco, la Fundación Asmoz de Eusko Ikaskuntza, la VPL del Gobierno Vasco e IVAP, las Diputaciones Forales de Vizcaya y Guipúzcoa, Grupo Voice S.L., Robotiker, UZEI y Vicomtech.

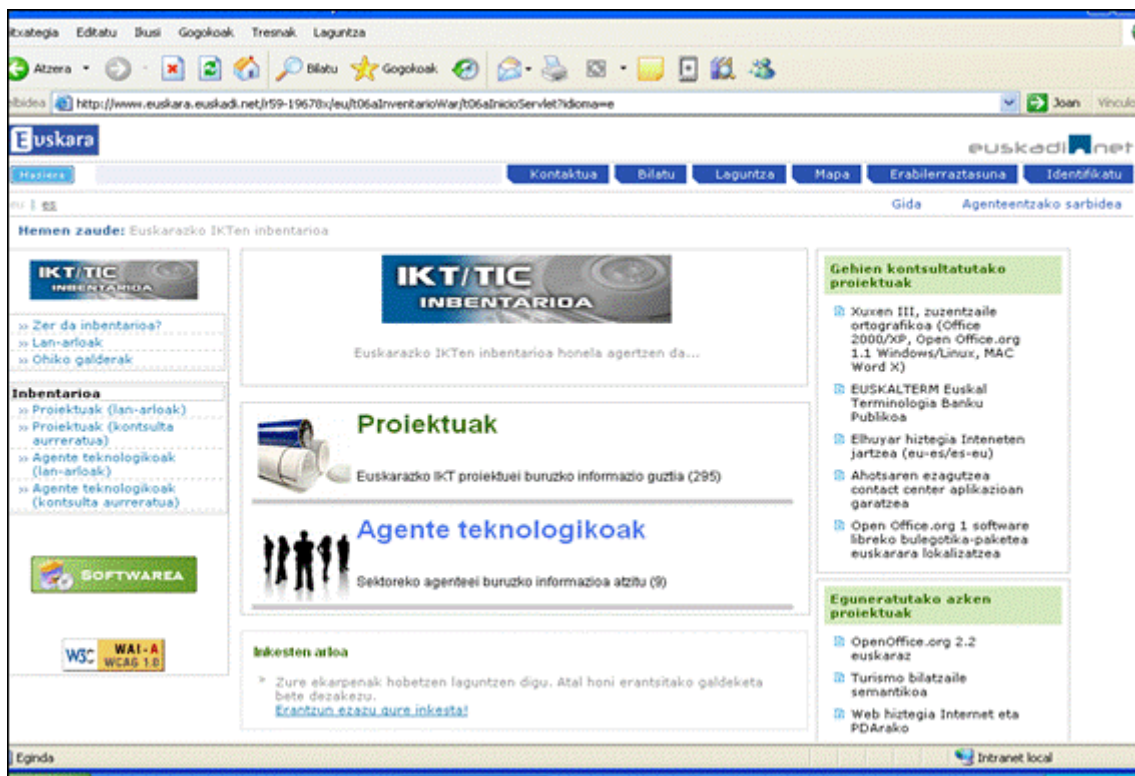
Sin embargo, con el fin de mantener el Inventario actualizado y vivo, la VPL desarrolló una aplicación a través de Internet que se encuentra disponible para todos los usuarios: por una parte, se ofrece un servicio de consulta general para buscar información sobre cualquier proyecto o producto, y por otra, los agentes tecnológicos tienen acceso directo para poder actualizar sus proyectos o insertar nuevos proyectos. Para poder acceder al Inventario como agente tecnológico es necesaria la tarjeta IZENPE (de los ciudadanos) y, además, hay que darse de alta en el sistema. Las instrucciones están detalladas en la misma aplicación. Además de la aplicación de Internet, los usuarios cuentan con la aplicación de intranet, cuya función es que la VLP gestione el inventario; entre otras

operaciones, sirve para validar la información recibida. De este modo se asegura que la información sea homogénea y correcta. Veamos la estructura del Inventario de las TIC :

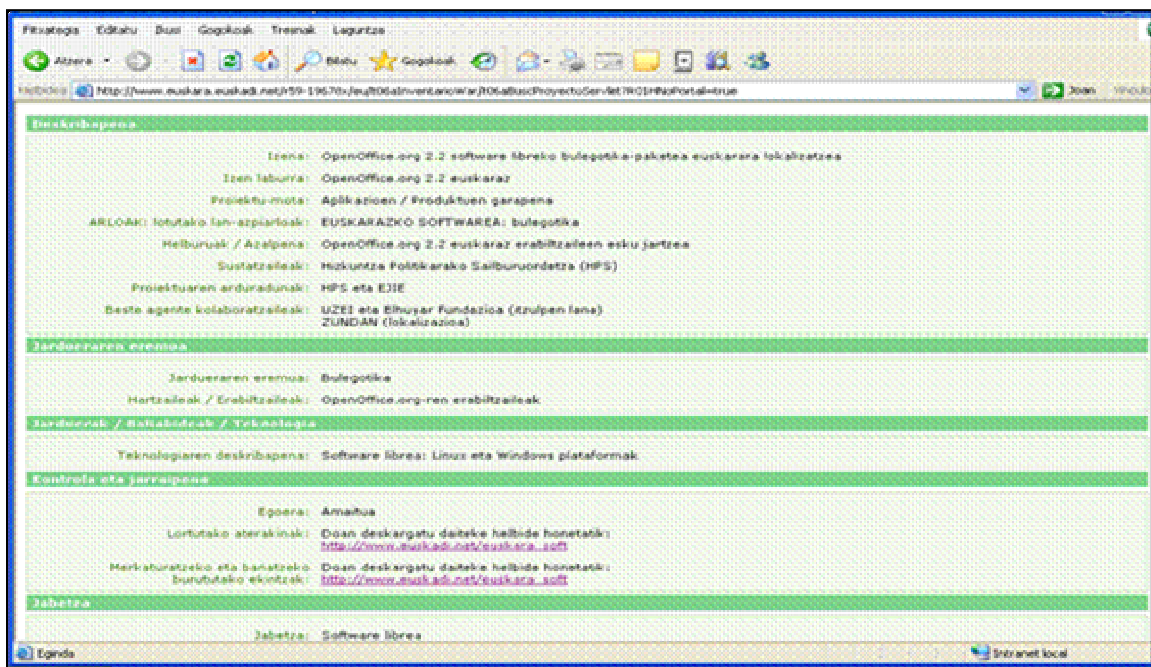
PRESENCIA DEL EUSKERA EN LAS NUEVAS TECNOLOGÍAS: Herramienta Inventario de las TIC en euskera



El inventario de las Tecnologías de la Información y de la Comunicación se encuentra en la siguiente dirección: www.euskara.euskadi.net/ikt_inventarioa. Está disponible para todos los usuarios en general, y los agentes tecnológicos tienen la posibilidad de insertar información sobre sus productos, como se ha indicado anteriormente:

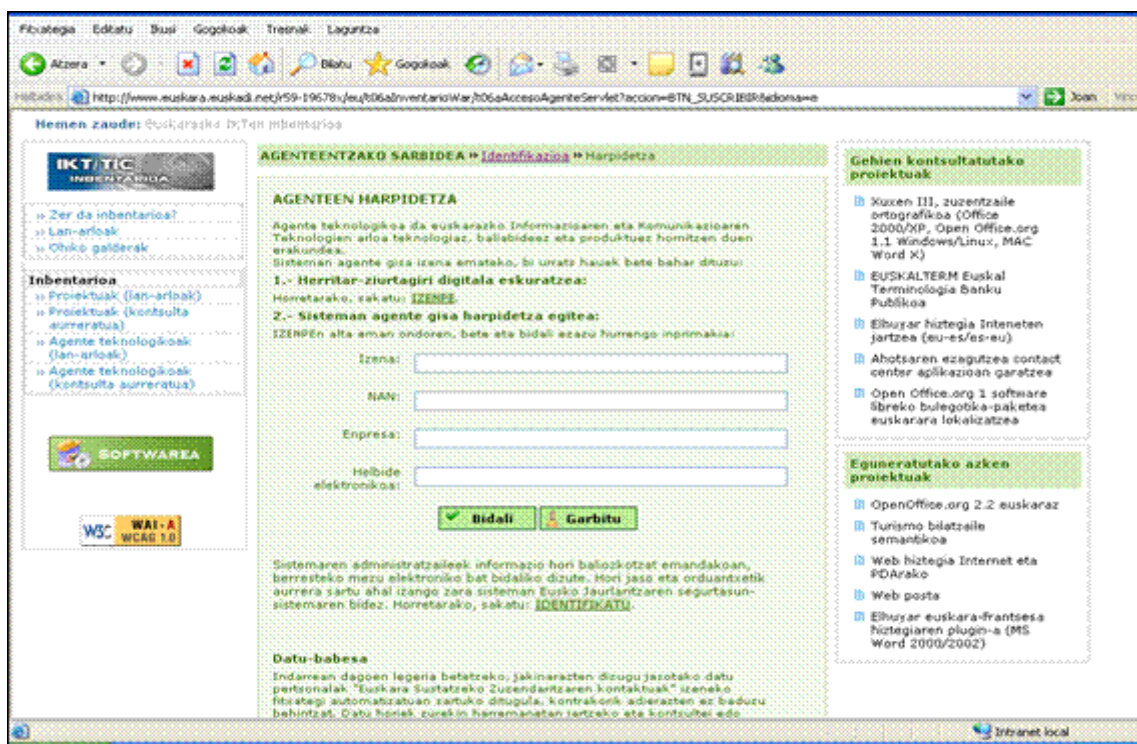


Si el usuario realiza una consulta sobre un proyecto en concreto (por ejemplo para OpenOffice.org), obtendrá la siguiente información:



El Inventario de las TIC se mantiene vivo sin duda gracias a las aportaciones de los agentes tecnológicos. Los consumidores de herramientas en euskera deben saber en todo momento qué se ha desarrollado en euskera, quiénes son los autores y cómo pueden conseguir información sobre ese producto o servicio. En definitiva, el objetivo es

promover e incrementar el mercado de las TIC en euskera. Como hemos mencionado, para utilizar el Inventario como agente tecnológico es totalmente necesario identificarse de forma segura (la tarjeta IZENPE de los ciudadanos) y estar dado de alta en el sistema:



La información que los agentes tecnológicos deben aportar sobre cada proyecto no se ofrece en su totalidad en Internet, ya que algunos campos sólo se utilizan para la gestión interna.

En este momento, el Inventario de las TIC recoge el siguiente número de proyectos/productos:

- Recursos y herramientas lingüísticas: 71
- Formación / e-learning: 8
- Herramientas de ayuda a la traducción: 15
- Sistemas interactivos avanzados: 19
- Gestión de la información y el conocimiento: 24
- Software en euskera: 199
- Industria de contenidos (crear contenidos digitales en euskera): 9
- Inclusión electrónica: 5
- Otros: 4

Debido a que existe la posibilidad de clasificar los proyectos/productos en más de un campo, hay varios productos/servicios que se repiten.

3. Comisión de las TIC del Consejo Asesor del Euskera (CAE)

El DECRETO 176/2007 del 16 de octubre sobre El Consejo Asesor del Euskera (BOPV, nº 204) establece lo siguiente: "El Pleno del Consejo Asesor del Euskera y el Consejero o Consejera de Cultura pueden decidir crear comisiones especiales para

realizar trabajos o informes sobre materias concretas. Para ello se requerirá el Acuerdo del Pleno del Consejo o la Orden del Consejero o Consejera del Cultura".

A continuación, la ORDEN del 30 de octubre de 2007 (BOPV, nº 235) de la Consejera de Cultura creó Comisiones Especiales del Consejo Asesor del Euskera. Entre otras, creó la comisión especial denominada "Tecnologías de la Información y de la Comunicación". Las tareas que la ORDEN encomienda a la Comisión son las siguientes:

"El objetivo de esta comisión especial es, entre otros, hacer propuestas para el adecuado desarrollo de la línea de actuación n.º 23 del documento Futuro de la política lingüística, que recoge el Proyecto 2005-2009 de la Viceconsejería de Política Lingüística. Sin duda es necesario desarrollar nuevos recursos, instrumentos y aplicaciones en euskera para las nuevas tecnologías de la información y la comunicación, así como incrementar la presencia del euskera en Internet, creando contenidos digitales. En este marco se considera de todo punto necesario analizar y radiografiar la presencia actual del euskera en ese ámbito y los recursos en euskera ya disponibles en estas nuevas tecnologías, así como definir el camino y las estrategias a seguir en el futuro.

La Comisión de las TIC del CAE está en marcha, y ha aprobado el Plan de Actuación de 2008. El punto de partida de la Comisión de las TIC del CAE es el Inventario de las TIC, dado que la información del Inventario es muy valiosa para analizar dónde pueden estar las carencias de las TIC en euskera y, como consecuencia, para definir las prioridades en este ámbito de trabajo. En el Plan de Actuación se recogen las siguientes prioridades:

- Realizar un estudio sobre INTERNET proponiendo las siguientes tareas: evaluar el uso del euskera (cuántos sitios web hay en euskera, la blogosfera vasca, número de personas que navegan en euskera), las herramientas de Internet en euskera (los grupos, productos, etc. que las desarrollan), los contenidos en euskera de la red, iniciativas para los jóvenes...
- Recursos y herramientas lingüísticas: definir los proyectos para la creación de dos corpus: el corpus de referencia del euskera del siglo XXI y el corpus bilingüe etiquetado para ser utilizado en traducción automática.
- Analizar las necesidades y proponer proyectos en los siguientes ámbitos: sistemas interactivos avanzados (tecnologías de voz, sistema para responder preguntas...); software en euskera (localizar al euskera las aplicaciones del mundo empresarial, software libre y/o propietario); gestión de la información y del conocimiento; e inclusión electrónica (aplicaciones para las personas con discapacidades).

4. ¿Y en el futuro?

Las grandes lenguas están desarrollando e impulsando las Tecnologías de la Información y de la Comunicación principalmente por dos tipos de razones:

- Económicas: el objetivo es invertir en tecnología, para después reducir el coste. Los usuarios de dichas tecnologías saben que todavía no son perfectas y que en algunos casos los resultados no son óptimos (por ejemplo, en el caso de la traducción entre dos lenguas muy diferentes). En cambio, el concepto de calidad está cambiando, y cada vez resulta más imprescindible entender el concepto de

calidad en base a otros dos conceptos: "cantidad, calidad y precio" (*quantity, quality, price*). La Unión Europea sigue este criterio plenamente, ya que con el incremento del número de lenguas oficiales los costes han aumentado significativamente.

- Plurilingüismo (*multilingualism*): la información y la comunicación son cada vez más complejas. Las prácticas que utilizan solamente una lengua o dos en las tecnologías lingüísticas quedarán marginadas definitivamente.

Y, si miramos hacia el futuro, ¿cuáles serán los campos de investigación y negocio en cuanto a las TIC se refiere? Las empresas y los centros de investigación están focalizando sus previsiones en los siguientes campos:

Tecnologías de la voz

- Teléfonos móviles: las empresas consideran que el negocio más importante en los próximos años estará en los teléfonos móviles (*key technology area*). Según las previsiones, el número de teléfonos móviles aumentará enormemente de aquí en adelante. Por lo tanto, hay que desarrollar sistemas para ofrecer servicios telefónicos en varios idiomas (información sobre los negocios, datos sobre el estado de las ciudades, direcciones, compras, reservas, información sobre el tiempo, programas de televisión, deportes, juegos...). Estos sistemas tendrán influencia en gran parte de la población mundial. Si comparamos el teléfono con Internet, a la hora de desarrollar aplicaciones, el teléfono ofrece más información: el género de la persona (femenino/masculino), edad, nivel de educación, raza... Por ejemplo, se dice que los resultados son muy diferentes cuando se utilizan con hombres o con mujeres. Por lo tanto, la clave en estos sistemas será realizar interfaces adecuadas y multimodales. Además, estos servicios vía teléfono serán muy útiles, por ejemplo, para las personas invidentes.
- Comprensión del lenguaje hablado (*spoken language understanding*), y, además, en un circuito de comunicación multilingüe.
- Reconocer la voz o el habla: controlar la voz, añadirle emociones, desarrollar sistemas interactivos...
- Traducción de habla a habla (*Speech-to-Speech Translation*): utilización simultánea de la tecnología de la voz y la de la tecnología de la traducción automática.

Traducción automática (Machine Translation MT)

La práctica confirma que los resultados de las traducciones realizadas entre idiomas muy similares son muy buenos (95%); en cambio, entre lenguas que son muy dispares, los resultados no son buenos por el momento (como mucho un 65%). En cualquier caso, como se ha explicado anteriormente, el concepto de calidad está cambiando, y ahora hay que medirlo junto a otros parámetros, es decir, junto con la cantidad y el precio. El consumidor de la traducción automática no busca la perfección. Su objetivo es conseguir una información general sobre un texto que le resulta incomprensible. Por ello, podríamos decir que el usuario prefiere que la información sea imperfecta, antes de no disponer de ninguna información. Por supuesto, hablamos de información general. Resulta evidente que en otros contextos, por ejemplo en el caso de textos oficiales o similares, se debe utilizar este tipo de sistemas con mucha precaución, y antes de

entregar el texto al cliente final resulta necesaria una corrección humana. Por lo tanto, junto con la utilización controlada de estos sistemas, el planteamiento es que hay que mejorar el sistema, es decir, hay que realizar un control de calidad. Ha de medirse también el tiempo de respuesta. Es cierto que con Internet la "demanda de una traducción rápida y barata" está aumentando.

Se están utilizando sistemas de evaluación (por ejemplo BLEU) para evaluar los sistemas de traducción automática.

Recuperar información (*Information retrieval*)

Son herramientas para buscar información relevante de manera rápida y eficaz. Para ello, se cruzan diferentes lenguas, es decir, se pueden realizar búsquedas plurilingües (*cross-lingual retrieval*). En este campo se sitúan también las bibliotecas digitales, esto es, sistemas que utilizando idiomas diferentes ponen a disposición de los usuarios documentos importantes y otros tesoros de todo el mundo. Esta tecnología utiliza el sistema pregunta-respuesta del mismo modo que ontologías y tesauros. Relaciona, clasifica y resume información. En definitiva, se trata de la tecnología de la web semántica: toma en cuenta conceptos, entidades y acontecimientos, realiza análisis de la estructura y coloca etiquetas semánticas.

Internet

Recientemente, alguien dijo que Internet es la nueva "ágora". Es decir, el eje de la era de la Información y de la Comunicación se encuentra en Internet. En los últimos años hemos observado una rápida evolución de la Red. Este último año, hemos pasado de la web social (2.0) a la web semántica (3.0). Cada vez hay más contenidos en Internet, y en más idiomas. Por ello, se necesitan sistemas para gestionar estos contenidos tan complejos. Y para que estos sistemas puedan aplicarse en los nuevos contextos, es necesario un lenguaje, un aprendizaje a través de la máquina y una tecnología para representar el conocimiento. Los sistemas se basan en la semántica, es decir, en los conceptos, en las relaciones entre los conceptos, y por consiguiente en las ontologías y en los tesauros. En la Web semántica también se utiliza el sistema pregunta-respuesta. Hay que resaltar que la diferenciación entre el contenido y el servicio está desapareciendo en Internet.

Pero junto con las **tecnologías del lenguaje** (*Language Technologies LT*), se deben ofrecer **recursos lingüísticos** (*Language Resources LR*). De hecho, los recursos lingüísticos resultan necesarios para el desarrollo de las tecnologías del lenguaje. Los recursos lingüísticos cobrarán cada vez más importancia en todos los idiomas, dado que todos serán necesarios para implementar servicios plurilingües. Por ello, los corpus (monolingües y plurilingües, paralelos o alineados), los recursos para las tecnologías de la voz, etc. se cotizarán al alza, especialmente para las empresas que quieran comercializar productos.

La organización europea ELRA (*European Language Resources Association*) ofrece recursos en todos los idiomas: www.elra.info.

En cuanto a los recursos lingüísticos se refiere, la Unión Europea muestra su preocupación por los formatos y los estándares, y está considerando establecer un modelo europeo para armonizarlos en los próximos años, puede que mediante una directiva. El objetivo sería establecer los campos prioritarios, siempre teniendo en cuenta los de más interés, y parece ser que no sólo se tendrían en cuenta las lenguas oficiales de la Unión Europea. De este modo, además de consolidar los métodos, se

realizaría un *roadmap*, y, a su vez, se establecerían los protocolos de evaluación. Recientemente, la Unión Europea ha liberado su corpus de traducción (memorias de traducción), que consta de 6 millones de documentos, para que cualquier persona pueda hacer uso de ellos.

En lo referente a las tecnologías lingüísticas, las tecnologías existen, pero todavía falta una infraestructura donde poder reunir las todas.

5. Los desafíos del euskera

El euskera necesita un desarrollo tecnolingüístico. Las lenguas que no se suban a este tren no tendrán futuro. Es evidente que el mercado requerirá este tipo de productos y que los consumirá en el idioma en el que se comercialicen. La estrategia consistirá en establecer prioridades, teniendo en cuenta en todo momento los productos que se introducirán en nuestro mercado en un futuro próximo. De este modo, el objetivo sería que estos productos estuvieran disponibles también en euskera. Se han trabajado y se están trabajando varios ámbitos, pero es necesario seguir trabajando. Como hemos señalado, la clave estará en Internet (y también en los teléfonos móviles), y, por lo tanto, está claro que hay que crear contenidos en euskera. Ésa debe ser la prioridad, porque la tecnología de la web semántica exige un gran volumen de contenidos, ya que se debe tener en cuenta que esta tecnología se basa en sitios web plurilingües.

6. Dos direcciones interesantes sobre tecnologías lingüísticas

<http://www.lt-world.org/> *Language Technologie world* (Alemania)

<http://www.lang-tech.org/> *LangTech is the european forum dedicated to communities and organisations involved in the development, deployment and exploitation of Language and Speech technologies.*

7. Bibliografía

- Aizpurua, Xabier; Diaz de Lezana, Araceli: "El euskera y las nuevas tecnologías de la información y de la comunicación: Programa de Industrias de la Lengua", in Segundo Congreso Europeo de Planificación Lingüística, Andorra la Vella, 2001.
- Diaz de Lezana, Araceli: "Euskara eta garapen teknolinguistikoa", www.erabili.com/zer_berri/muinetik/1057933133, 2003.
- Diaz de Lezana, Araceli: "El desarrollo tecnolingüístico del euskera", en el Congreso Linguapax, Barcelona 2004.
- Futuro de la Política Lingüística. Proyecto 2005-2009 de la Viceconsejería de Política Lingüística Servicio Central de Publicaciones del Gobierno Vasco, Vitoria-Gasteiz, 2005, ISBN 84-457-2407-X.
- Langtech 2008, Congreso sobre las Tecnologías Lingüísticas, Roma, 2008.
- Wikipedia, http://en.wikipedia.org/wiki/Main_Page