

LA TRADUCCIÓN AUTOMÁTICA.

ESTADO DE LA CUESTIÓN.

Y

EVALUACIÓN DE LOS BUSCADORES.

**José Antonio Camacho Espinosa
María González Targhetta**

Facultad de Documentación - Universidad de Alcalá

Abril - 2002

LA TRADUCCIÓN AUTOMÁTICA. ESTADO DE LA CUESTIÓN.

José Antonio Camacho Espinosa y María González Targhetta¹

Facultad de Documentación - Universidad de Alcalá

RESUMEN

Se aborda el estado de la cuestión en el ámbito de la T.A. a través de la revisión bibliográfica de una pequeña muestra de artículos que se han considerado relevantes, tanto de autores españoles como extranjeros.

PALABRA CLAVE

Traducción Automática, Traductores, Arquitectura de la Traducción Automática, EUROTRA. TOPÓNIMOS: Europa.

ABSTRACT

The state of the question of the M.T. is approached through the bibliographical revision of a small sample of articles that they have been considered outstanding, so much of Spanish authors as extranejeros.

KEYWORDS

Machine translation, Translators, Machine Translation Architecture, EUROTRA. TOPONYMS: Europe.

A. Introducción

Presentamos en este trabajo la situación actual de la Traducción Automática (T.A.), para lo cual hemos hecho una selección de la bibliografía accesible en la red. En primer lugar se definen los fundamentos de la T.A., tanto desde una óptica lingüística como tecnológica y pragmática, para pasar a continuación a exponer brevemente la evolución histórica sufrida por estos sistemas en sus escasas seis décadas de vida. A continuación se describen los tipos básicos de T.A. para hacer un análisis más

¹ Correo electrónico: jacamach@worldonline.es y mariagt13@hotmail.com

pormenorizado de la arquitectura del modelo más extendido hasta la actualidad: el basado en reglas, de tipo indirecto. Se aborda el conflicto que se ha suscitado en ciertos círculos entre los traductores profesionales y los creadores o partidarios de la T.A., para continuar con un estudio detallado de los sistemas de T.A. desarrollados en la Comunidad Europea. Concluimos el trabajo con las perspectivas de futuro de estas herramientas cuyo auge crece de día en día.

B. Metodología

En primer lugar examinamos el temario propuesto por el profesor de la asignatura e hicimos una rápida revisión bibliográfica del mismo para determinar el tema de trabajo. Una vez elegido el de la T.A. procedimos a la búsqueda documental en la red a través de buscadores de propósito general, directorios y bases de datos o catálogos especializados. Vistos los documentos a los que podíamos acceder a texto completo y hecha la correspondiente selección, procedimos a su lectura para determinar las líneas de investigación o aspectos más sobresalientes de la materia y presentar, así, el estado de la cuestión. Por último se extraen unas conclusiones que tratan de dar una visión de conjunto del panorama de esta materia.

C. Resultados

1. FUNDAMENTOS DE LA T.A.

La traducción automática es tan antigua como la propia informática. Esta disciplina ha contribuido de manera determinante al desarrollo de la tecnología lingüística y de todas aquellas herramientas que tienen que ver con el tratamiento de la lengua en sus diversas facetas.

La información que circula actualmente entre empresas, instituciones y particulares de unos países a otros y, por tanto, entre personas con distinta lengua, es de cientos de millones de páginas diarias. *Hurbert Murria* cifró en 20 millones el número de palabras de información técnica que se generaban en el mundo cada día (Abaitua 2002a). En algunos casos los costes económicos para empresas e instituciones son de dimensiones tales que, como el caso de una empresa de automóviles, se llegan a traducir más de 300.000 págs. anuales de documentación; o la propia CEE, cuyos gastos de traducción e interpretación suponen casi el 50% del presupuesto de administración.

Tales demandas de traducción son las que han llevado –a pesar de la crisis atravesada en los años 80- a la investigación y creación de múltiples soluciones automatizadas.

Las necesidades básicas por las que un usuario precisa de la traducción es:

- 1) Para disponer de un texto que publicar.
- 2) Para hacerse una idea rápida del contenido de un texto.
- 3) Para acceder rápidamente a un texto de carácter urgente y, normalmente, de uso interno o profesional.
- 4) Para redactar en una lengua distinta de la propia. En la actualidad, la T.A. prácticamente es el único procedimiento capaz de dar respuesta a las tres últimas.

Por otra parte, nos encontramos con otros dos casos en los que un usuario precisa de la traducción de un idioma a otro:

- 1) Cuando se produce una comunicación oral entre dos o más personas que conocen distintos idiomas.
- 2) La traducción de programas o recursos informáticos. En el primer caso hablaremos, no de traducción, sino de *interpretación*. Este campo está destinado casi en exclusividad –por el momento- a los traductores humanos. En el segundo caso se aplica el término *localización* y viene a indicar cómo un recurso informático es adaptado tanto lingüística como culturalmente a las necesidades de una población o grupo humano. Aquí la T.A. tiene una labor trascendental, si bien con la participación correctora del hombre.

Hay que distinguir básicamente dos tipos de usuarios de los sistemas de T.A.: los que podríamos llamar usuarios finales técnicos, profesionales o administradores, cuya necesidad se centra básicamente en conocer el contenido de un documento en su conjunto o de una parte del mismo. Y los traductores, personas que utilizan las máquinas como ayuda para facilitar, y mejorar en rapidez y en calidad su trabajo. (Equipo de Traducción, 2001)

Por lo que hace a la tipología de textos a los que se puede enfrentar un proceso de traducción, podemos hacer un análisis desde distintas dimensiones que nos van a determinar de qué texto se trata (Abaitua, 2002a):

- 1) **Tipología del texto:** nos referimos al registro en el que se encuentra el texto (categoría social, profesional, técnica...); el estilo del mismo (directo-indirecto; formal-informal; etc.); la función pragmática con la que se creó el mismo o género literario (novela, ensayo, teatro...); y dentro de cada género, podemos encontrar distintos tipos (dentro del lenguaje periodístico, por ej., reportaje, crónica, artículo de opinión, etc.).

- 2) **Densidad léxica:** aquí tomaremos el término *granulidad* como indicativo del tamaño de las unidades léxicas en las que se puede fragmentar un texto. La densidad es inversamente proporcional a la *granulidad*. Los textos con una densidad más alta son los de tipo técnico o especializado.
- 3) **Distancia lingüística y cultural:** un aspecto poco tenido en cuenta en la mayor parte de los programas creados hasta ahora ha sido la distancia cultural que, unida a la lingüística, ha influido de forma decisiva en la baja calidad de los resultados.
- 4) **Desarrollo y recursos elaborados:** para poder crear herramientas informáticas como los T.A. es preciso disponer de recursos previos como diccionarios, bases de datos léxicas, análisis morfológico, sintáctico y semántico de la lengua, y, sobre todo, un corpus terminológico suficientemente amplio. Asimismo, hay que llevar a cabo un proceso de normalización en el uso del lenguaje para que un documento determinado, en cualquier lengua, cumpla con ciertos requisitos que lo hagan fácilmente traducible.
- 5) **El medio y el modo de la comunicación:** la oral requiere unos procedimientos muy distintos a la escrita.
- 6) **Los fines de la traducción:** hay una traducción de calidad, básicamente con el propósito de la posterior publicación, y una traducción rápida de baja calidad con el fin de dar una respuesta inmediata a las necesidades de un usuario.

En general podemos afirmar que los sistemas de T.A. obtienen mejores resultados cuando están diseñados con un propósito específico -ciñéndose a lenguajes especializados en un área concreta del conocimiento- que aquellos que han sido creados con propósito general.

Según Martín Kay, citado por Abaitua (2002b), un sistema de T.A. que quiera alcanzar sus objetivos habrá de cumplir las siguientes premisas:

- Mantener un equilibrio entre el traductor humano y el equipo (interactividad).
- La máquina debe ayudar al traductor –o al usuario- con todos aquellos instrumentos de los que dispone la informática: repertorios terminológicos, bases de datos de ejemplos, modelos, etc.
- La máquina debe permitir la integración de los procesos de redacción y traducción.

2. DESARROLLO HISTÓRICO

Si bien el autor más destacado en el campo de la T.A. es el británico John Hutchins –a quien nos referiremos más adelante- a la hora de revisar el desarrollo histórico de esta disciplina nos

vamos a ceñir al autor Johnatan Slocum, citado por Abaitua (2002a). El estudio histórico que hace este autor lo divide en seis periodos, que ajusta a cada una de las décadas del siglo XX a partir de 1940.

1. Los primeros desarrollos se realizaron con el famoso ordenador ENAC en 1946. Entre los investigadores pioneros hay que destacar Warren Weaver, de la Fundación Rockefeller. Tanto los ordenadores como las técnicas empleadas eran aún muy rudimentarias.
2. En 1951 el M.I.T. puso a trabajar en un proyecto a Yehoshua Bar-Hiller y un año más tarde se organizó el primer simposio de la T.A.. La primera demostración pública se llevó a cabo en 1954 en la Universidad de Georgetown, con la participación de la empresa IBM y del investigador Leon Dostert. Fue la época de la euforia de la FAHQT o *traducción totalmente automatizada de alta calidad*. De esta época son dos de los proyectos que mayor incidencia han tenido: SYSTRAN Y METAL.
3. La década de los 60 trajo el contrapunto a la anterior cuando en 1964 el *National Research Council* creó un comité (ALPAC) para evaluar los sistemas creados hasta ese momento y sus resultados. Las conclusiones fueron desastrosas, dado los objetivos tan exigentes que se habían planteado en la década anterior. Pese a ello fue una época de grandes avances, sobre todo en el campo de la Lingüística, con aportaciones de la categoría de las de Noan Chomsky.
4. En 1976 un grupo de investigadores de la Universidad de Montreal presentan el sistema MÉTÉO que traducía los partes meteorológicos del inglés al francés. Uno de los sistemas más sobresalientes en la breve historia de los T.A. por su calidad e idoneidad. Es el momento en que la Comunidad Europea adquiere los derechos de SYSTRANS para, poco después, poner en marcha el proyecto EUROTRA –del que hablaremos más adelante.
5. En la década siguiente se producen dos acontecimientos destacables y no para bien: la C.E. decide cancelar la financiación del proyecto EUROTRA y la empresa PHILIPS da por terminado uno de los proyectos más prestigiosos: ROSETA. Hacen su aparición un conjunto de productos con un nuevo diseño creados por IBM, TRADOS, STAR Y ATRIL.
6. En los 90, Internet irrumpe con fuerza. Es el “leitmotiv” de todos los desarrollos actuales. La globalización de los mercados y de la información obligan a centrar la investigación tanto en la “localización” de programas como en la creación de documentos en formato electrónico en varios idiomas a la vez.

3. TIPOS Y ARQUITECTURA DE LOS SISTEMAS DE T.A.

La T.A. ha tenido hasta el presente dos líneas de investigación y desarrollo bien diferenciadas: basada en reglas y basada en analogías.

Las primeras –también cronológicamente- tienen su origen en la Lingüística Generativa y en la Inteligencia Artificial, sobre todo a partir de 1970. Los investigadores que seguían estas tendencias consideraban la traducción como un problema de equivalencia semántica, en la creencia de que todas las lenguas del mundo comparten una misma subestructura lógica.

Entre estos sistemas podemos diferenciar los métodos directos, que hacen una traducción palabra por palabra -que fueron abandonados hace tiempo- y los métodos indirectos. Entre estos últimos hay que destacar los que se apoyan en la técnica de transferencia, que realizan el proceso en tres fases:

- 1) **Análisis** de la lengua de partida: para intentar resolver casos de homografía y hacer comprobaciones estructurales, intentando determinar las relaciones básicas entre palabras. El objetivo es extraer de la lengua de partida el mayor número de datos posible de carácter sintáctico, semántico, etc.
- 2) **Transferencia**: es cuando realmente el sistema hace la traducción, pero en su forma canónica, bien a nivel léxico, a nivel sintáctico o a nivel semántico.
- 3) **Síntesis** a la lengua de llegada: En realidad se trata del proceso simétrico al de análisis para dar forma final legible y entendible al texto ya traducido. (Equipo de Traducción, 2001; Canals et all.)

El drástico abaratamiento de los microprocesadores y de las unidades de almacenamiento, junto a la proliferación de los textos electrónicos y de Internet llevó a los investigadores a adentrarse en líneas de trabajo basadas en estudios probabilísticos. Autores como Nord, que hacen hincapié en los aspectos estilísticos y pragmáticos de la lengua, o Hatim y Mason que destacan el campo de la semiótica, obligan a un cambio de rumbo a los informáticos. Con estas tendencias se han creado en la última década varias soluciones: unas basadas en **métodos estadísticos**, como el EGYPT, del grupo ISI de la University of Southern California; el **método basado en ejemplos**, extrayendo oraciones o sintagmas de un corpus de textos bilingües, previamente alineado (proyecto de Makoto Nagao en la Universidad de Kyoto o el DLT en Holanda); y una tercera, conocida como **memoria de traducción** y que consiste en el almacenamiento de traducciones realizadas por personas y reutilizarlas posteriormente en textos similares. Aquí hay que destacar la empresa alemana TRADOS o la española ATRIL. (Abaitua, 2002a)

La creación de un sistema de T.A. es un trabajo de equipo que exige la más estrecha colaboración entre lingüistas y traductores, encargados de seleccionar los datos, los conceptos y las reglas que rigen la comunicación humana, y los informáticos que escriben los programas compiladores, encargados de generar automáticamente los subprogramas del sistema.

Los programas más desarrollados hasta hace pocos años eran los T.A. por transferencia. En este modelo se pueden diferenciar cinco tareas que se asignan a otros tantos subprogramas (Canals et al.):

1. Análisis morfológico de las palabras del texto original usando: un diccionario de lemas, una descripción de los paradigmas de flexión y la relación entre lemas y paradigmas. Esta tarea se asigna al **compilador de analizadores morfológicos**.
2. Desambiguación categorial de los homógrafos, usando información estadística sobre la aparición conjunta de términos, que realiza el **desambiguador de categorías léxicas**.
3. Consulta del diccionario bilingüe para hacer corresponder los lemas de la lengua de origen y los de la de destino. Realizado por el **compilador de diccionarios bilingües**.
4. Detección y tratamiento de secuencias de palabras usando las reglas adecuadas, que corre a cargo del **compilador de subprograma de transferencia**.
5. Generación de las formas superficiales de la lengua de destino, gracias al **compilador de generadores morfológicos**.

4. LA T.A. Y LA TRADUCCIÓN HUMANA. ENFRENTADAS O COMPLEMENTARIAS?

Traducción automática y traducción humana en ningún momento –al menos por ahora- se van a enfrentar por cuanto una y otra cumplen funciones, en unos casos distintas y en otros complementarias. En este contexto es importante distinguir tres modos de abordar la traducción con la ayuda del ordenador:

- 1) La T.A. con el fin de tratar el proceso en su totalidad, cuyo resultado invariablemente es necesario revisar.
- 2) Los sistemas de T.A. como apoyo al profesional de la traducción.

- 3) Los sistemas de T.A. para el usuario ocasional, que sólo proporciona traducciones muy rudimentarias, con el objetivo último de que el usuario comprenda el contenido del documento.

La mayoría de los expertos en la materia están de acuerdo en que los primeros productos aparecidos en el mercado no fueron herramientas útiles ni adecuadas para el traductor profesional. En la década de los 90 se han creado los llamados **puestos de trabajo de traducción** que ofrecen a los profesionales una serie de herramientas para hacer un trabajo más productivo: acceso a diccionarios, bases de datos terminológicas, glosarios, medios para la entrada y tratamiento de textos (por ej. a través del OCR) e incluso acceso a traducciones ya existentes recogidas también en bases de datos, quizás la ayuda más apreciada por el traductor (Hutchins, 2001b). Uno de los recursos más eficaces integrado en estos "puestos de traducción" es el de **localisation** de software que permite adaptar la documentación electrónica los programas informáticos a las particularidades del contexto de cada país o comunidad a la que va destinado.

5. LA T.A. EN EUROPA

En 1976² la Comisión Europea adquirió los derechos de SYSTRAN, creado en EE. UU. a finales de los años 50 por Peter Toma. Este sería el primer paso del proyecto EUROTRA. El primer desarrollo llevado a cabo fue el par inglés-francés y desde entonces se han ido añadiendo paulatinamente nuevas lenguas, si bien quedan aún un gran número de pares por desarrollar. La calidad de cada par viene determinada por lo avanzado de su desarrollo y por la afinidad sintáctica y léxica de las lenguas que lo forman. Los mejores resultados se logran con los pares que incluyen inglés, francés, español e italiano (Equipo de Traducción, 2001)

Este sistema, alojado en el Centro de Datos de la Comisión, ha funcionado mediante:

- A. Programas: unos de carácter básicos, escritos en IBM Asembler y otros de tipo lingüístico, en SPL (*SYSTRAN Programming Language*). Desde 1997 se produjo la migración de los programas a Lenguaje C.
- B. Diccionarios: un diccionario básico constituido por unitérminos, otro de expresiones o reglas contextuales y, desde los años 90, se incluyeron las traducciones contenidas en EURODICAUTOM.

² Ver el capítulo 2.

El proyecto EUROTRA nació con un carácter exclusivamente universitario y ha sido el más ambicioso de cuantos se han desarrollado hasta el presente. Con él la Comunidad Económica Europea se propuso:

1. La construcción de un prototipo preindustrial de traducción automática para las lenguas oficiales de la Comunidad.
2. La formación de expertos en traducción por ordenador y otras áreas de la lingüística computacional.

El proyecto tenía como fin último conseguir traducciones de suficiente calidad sin la intervención humana antes o durante el proceso, aunque no se descartaba cierta intervención de postedición. Lo ambicioso de este proyecto era la creación de los 72 módulos correspondientes a las 9 lenguas que en su momento se hablaban en la C.E.E. y que se han visto incrementadas posteriormente con la incorporación de nuevos países. Una de las tareas más costosas tanto en recursos económicos como humanos ha sido -y es- la creación de los diccionarios y los repertorios terminológicos, labor a la que se dedican varios grupos de investigación de universidades tanto europeas como de fuera de la C.E. y entre los que destacan dos españoles, uno de la Universidad Autónoma de Madrid y otro de la Universidad Central de Barcelona. (Marcos Marín et al., 1988-89)

Dados los cambios tan notables que han supuesto en los últimos años las tecnologías de la información y la comunicación (TIC) y los procesos de globalización, la C.E. ha puesto un énfasis especial en las "industrias de la lengua" y en la tecnología necesaria para su desarrollo. Un proyecto que ha venido a sustituir a EUROTRA –si bien partiendo de sus logros- ha sido EURAMIS (*European Advanced Multilingual Information Systems*) que comenzó sus andanzas en 1994. El entorno incorpora: a) un gestor de memorias de traducción de TRADOS; b) un extractor de terminologías; c) el sistema de traducción SYSTRAN; d) un paquete de servicios lingüísticos. El diseño prevé que cualquier trabajador que solicite una traducción la envíe por correo electrónico al gestor del programa. Éste convierte el documento a SGML, siendo codificado con Unicode. Una vez procesado el texto por el sistema, se envía al equipo de traductores humanos quienes lo revisan y lo devuelven al cliente. (Abaitua, 2002b)

6. LA T.A. EN EL FUTURO

El futuro de la T.A. pasa por la normalización de la documentación escrita y tiene su campo de actuación más destacado en el lenguaje estático, por oposición al lenguaje dinámico que se encuentra en los textos espontáneos o creativos en el habla coloquial. La experiencia de estos dos últimos años, especialmente en ámbitos de gran consumo como la Comisión Europea o la Organización Panamericana de la Salud, nos hace ver que:

1. La cercanía entre los usuarios de las herramientas y sus creadores es garantía de que éstas se adaptan mejor a los objetivos perseguidos.
2. Las herramientas se deben integrar en el flujo documental, con soluciones de software para el tratamiento de formatos, tipologías de textos, caracteres, filtros, correo electrónico, etc.
3. El desarrollo de los diccionarios es una de las claves del éxito de los sistemas.

Se trata, en definitiva, de crear sistemas expertos en el tratamiento de textos, capaces de ofrecer al redactor modelos precisos de documentos, donde se combinen las hojas de estilo y las herramientas de traducción más avanzadas. (Abaitua, 2002b)

El énfasis de las investigaciones y desarrollos tecnológicos de futuro, según John Hutchins (2001a), está en la creación de herramientas que permitan la intercomunicación entre personas con diferentes idiomas, que puedan cubrir una amplia gama de necesidades, según la tipología de usuarios y los objetivos perseguidos por éstos. No obstante, no habrá que olvidar que ninguna máquina podrá emular la comunicación humana, con sus aspectos emocionales, de persuasión, afectivos... El criterio principal a la hora de evaluar la calidad de un sistema de T.A. seguirá siendo la utilidad, su capacidad para alcanzar los objetivos que se ha propuesto el usuario en un momento dado.

En las próximas décadas habrá que abordar (Hutchins, 2001a):

1. La ampliación del rango de idiomas tratados por los sistemas de T.A.
2. La ampliación, asimismo, del rango de estilos y registros que se vienen empleando.
3. Mejora en la calidad de la traducción de documentos de carácter personal.
4. Creación de nuevos instrumentos para la recuperación de información, revisión de documentos y resumen e indización de los mismos, aplicándoles la T.A.
5. Mayor calidad en las áreas donde la traducción humana no es imprescindible: traducción en tiempo real, en línea y traducción de baja calidad.

En definitiva, se trata de dar calidad al usuario y responder a las necesidades que van surgiendo cada día.

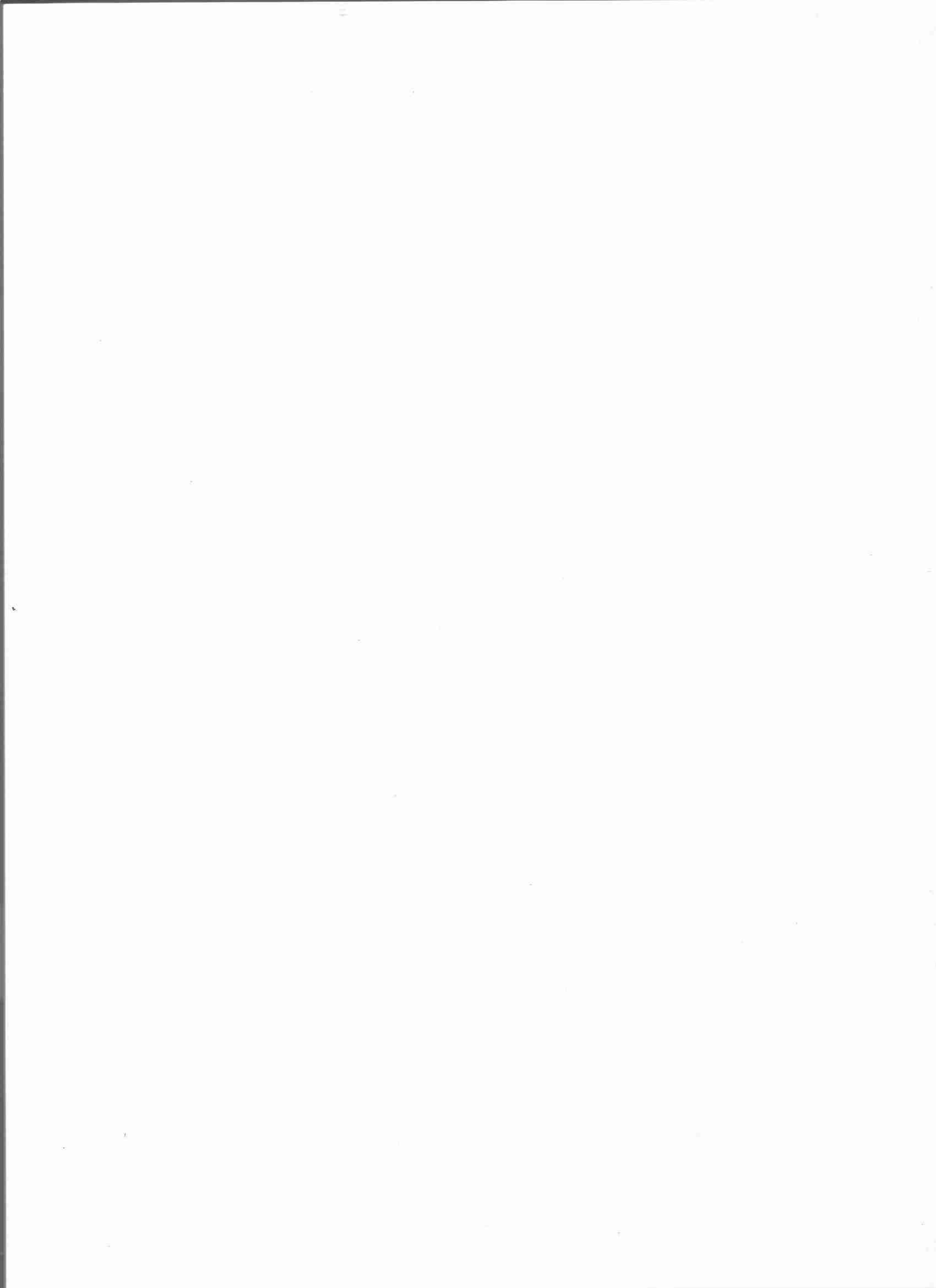
D. Conclusiones

Observamos, a través de la revisión bibliográfica realizada sobre la T.A., que:

1. Esta disciplina es prácticamente tan antigua como los propios ordenadores.
2. En las primeras décadas los investigadores se plantearon unos objetivos demasiado ambiciosos que trajeron un período de crisis tras el informe ALPAC.
3. Los logros alcanzados en la T.A. se han visto limitados al mantener durante años la tesis de la equivalencia lingüística de las lenguas y la creencia de que todas compartían una misma subestructura lingüística.
4. Nuevas líneas de investigación -que introducen elementos estilísticos, pragmáticos o semióticos de la lengua-, unidas al abaratamiento de los procesadores y de los recursos de almacenamiento de información, han permitido crear sistemas basados en modelos probabilísticos que están siendo más eficaces.
5. La T.A. obtiene mejores resultados en áreas restringidas -trabajando con corpus terminológicos especializados- que con implementaciones de propósito general.
6. En la última década la T.A. ha tenido una fuerte demanda con el desarrollo de Internet, la masiva creación de documentos electrónicos y la globalización de la información.
7. Los sistemas de T.A. tienden a ser cada vez más especializados, destacándose los llamados *puestos de trabajo de traducción* creados como herramienta de gran potencia para traductores profesionales.
8. La creación y paulatina ampliación de la C.E. ha sido uno de los factores que mayor peso han tenido en la puesta en marcha de proyectos de T.A. Los más destacados: EUROTRA Y EURAMIS.
9. El futuro camina hacia sistemas capaces de tratar un número de lenguas cada vez mayor, integrando múltiples herramientas capaces de hacer un tratamiento completo de un texto en el mismo momento en que se está creando.

BIBLIOGRAFÍA

- ABAITUA, J. (2002a) *Introducción a la traducción automática* [en línea] http://www.serv-inf.deusto.es/abaitua/konzeptu/ta/mt10h_es/ta10h-1es.htm [Consulta: 7 marzo 2002]
- ABAITUA, J. (2002b) *Traducción automática: Presente y Futuro* [en línea] <http://www.foreignword.com/es/Technology/art/Abaitua/default.htm> [Consulta: 7 marzo 2002]
- CANALS, R. et al *Herramientas para la construcción de sistemas de traducción automática: aplicación al par castellano-catalán* [en línea] www.torsimany.ua.es/docum/articad.pdf [Consulta: 7 marzo 2002]
- EQUIPO DE TRADUCCIÓN AUTOMÁTICA, SERVICIO DE TRADUCCIÓN DE LA COMISIÓN EUROPEA (2001) *SYSTRAN CE: el sistema de traducción de la Comisión Europea* [en línea] http://216.239.37.100/search?q=cache:JOVU8Fpq_2cC:europa.eu.int/comm/translation/en/eyl/mtsumes.pdf+systran+sistema+traduccion+automatica+comision+europa&hl=es&lr=lang_es [Consulta: 7 marzo 2002]
- HUTCHINS, J. (2001a) *Towards a new vision for MT* [En línea] VIII Conferencia sobre Traducción Automática <http://ourworld.compuserve.com/homepages/WJHutchins/MTS-2001.htm> [Consulta: 7 marzo 2002]
- HUTCHINS, J. (2001b) Machine translation and human translation: in competition or in complementation? [en línea] *International Journal of Translation*, nº 13, 2001 <http://ourworld.compuserve.com/homepages/WJHutchins/IJT.htm> [Consulta: 7 marzo 2002]
- MARCOS MARTÍN, F; MORENO SANDOVAL, A.; SÁNCHEZ LEÓN, F. (1988-89) El proyecto EUROTRA en el marco de la investigación sobre traducción por ordenador [en línea] *Telos*, nº 16 Diciembre-Febrero 1988-89 <http://www.quademsdigitals.net/articles%5CTELOS%5Ctelos16%5Ct16proyecto.htm> [Consulta: 7 marzo 2002]



EVALUACIÓN DE BUSCADORES

José Antonio Camacho Espinosa y María González Targhetta ¹

RESUMEN: Se trata de identificar las características principales de los Metabuscadores, Motores de búsqueda, Índices temáticos y Directorios más representativos así como su funcionamiento. Se planteará una ecuación de búsqueda que será lanzada a todos ellos y se evaluarán y compararán sus resultados.

PALABRAS CLAVE: Evaluación, búsqueda de información, metabuscador, motor de búsqueda, directorio temático, herramientas de búsqueda,

ABSTRACT: It is pretended to identify the meta-search engine, search engine, and directory features and proceedings. In each of them a query will be enter and the results will be analiced and evaluated.

KEYWORDS: Evaluation , search tools, information search, meta-search engine, search engine, directory.

I. INTRODUCCIÓN

En poco más de dos años Internet ha pasado a tener más de 2 billones de páginas web. Y esta cantidad va incrementándose de día en día. Se ve, por ello, imprescindible el uso de herramientas adecuadas que pongan en orden esa cantidad de información. Los Motores de búsqueda, Índices temáticos y Directorios han irrumpido, por tanto, como algo absolutamente necesario. Sin embargo, su proliferación es la segunda parte del problema, en el sentido de que se hace imprescindible ver qué es lo que se nos está ofreciendo, qué buscadores nos convienen en cada momento y para qué.

II. OBJETIVOS

Alcanzar un conocimiento global de las características de las distintas herramientas de búsqueda

Poner a prueba su funcionamiento

Comparar sus resultados de búsqueda

III. METODOLOGÍA

En primer lugar se diseñaron unas plantillas con las que organizar los datos de la evaluación de cada buscador. Posteriormente, tras la correspondiente consulta bibliográfica tanto de monografías y artículos de revista como de documentos en la red, se procedió a hacer un estudio teórico de cada una de las herramientas de búsqueda. Después, se redactó una ecuación de búsqueda sobre un tema de nuestro interés para la puesta en práctica de esos conocimientos teóricos, de manera que se pudiera introducir la misma ecuación –o similar- en cada una de las herramientas para posteriormente ir analizando y anotando los resultados de cada una.

¹ Correo electrónico: jacamach@worldonline.es y mariagt13@hotmail.com

IV. RESULTADOS

En cada uno de los cuadros donde se recoge la información de cada motor de búsqueda se han ido recogiendo datos siguiendo la guía o plantilla que se describe a continuación:

1. INTERFASE

a. Aspectos formales y gráficos

- **Consistencia, claridad y legibilidad:** homogeneidad y equilibrio en el colorido; sobriedad en el diseño; fondos claros; tamaño de tipografía legible; animaciones que no distraigan.
- **Estabilidad:** disposición estable de las distintas páginas de un mismo buscador; no hay cambios bruscos en la organización de las páginas, en los colores, en los símbolos..
- **Imágenes:** iconos intuitivos, limpieza de presentación, agilidad al cargarse la pantalla.
- **Enlaces visitados:** se diferencian de los no visitados con distinto color o tipografía.
- **Enlaces numerados:** se numeran los enlaces de forma continua.

b. Contenido

- **Autoría:** bien definida.
- **Servicios:** ofrece una relación de servicios de interés para el usuario. (citar)
- **Información general:** ofrece información de interés para el usuario (citar)
- **Sistematización:** presenta los resultados organizados de forma adecuada.
- **Información de enlaces:** presenta información suficiente de cada enlace antes de conectar con ellos.
- **Rutas:** se indica el camino recorrido hasta llegar a un sitio web (en el caso de directorios temáticos)
- **Palabras clave:** aparecen señaladas en cada enlace o en el resumen del mismo las palabras que identifican el contenido de la pregunta.

c. Aspectos técnicos. Navegación

- **Software:** utiliza software estándar y de fácil funcionamiento.
- **Frames:** evita su uso.
- **Ayuda:** ofrece ayuda al usuario durante la navegación a través de conexión con tesauros, listas temáticas, sugerencias....
- **Resultados:** permite elegir al usuario la presentación de resultados según distintos criterios.
- **Valoración:** valora los resultados obtenidos.
- **Menús claros:** los menús se presentan de forma clara y uniforme, facilitando la navegación.
- **Ampliación:** remite a otros enlaces alternativos (en caso de no obtener resultados) o complementarios (para ampliar la oferta de documentación al usuario).

2. BÚSQUEDA AVANZADA

- **Disponible:** ofrece las dos posibilidades de búsqueda simple y búsqueda avanzada.
 - **Sintaxis:** tipo de sintaxis de que dispone; cómo se presenta al usuario.
 - **Página de ayuda:** existe una página de ayuda al usuario para explicar cómo hacer la búsqueda avanzada.
 - **Ayuda de calidad:** la ayuda que se ofrece es sencilla, está bien explicada y es de calidad.
3. PRECISIÓN / PERTINENCIA
- **De la 1ª pantalla:** % de resultados pertinentes.
 - **De los 100 primeros resultados:** : % de resultados pertinentes.
4. RECALL
- Grado de recuperación de páginas pertinentes sobre el total de páginas pertinentes de la red. Para hacer este cálculo nos hemos basado en los resultados de la búsqueda realizada en el metabuscador Dogpile.
5. ARQUITECTURA Y DISEÑO
- Información acerca de los instrumentos técnicos, bases de datos, software, etc. Que definen la estructura interna de cada motor de búsqueda. Este apartado no se ha tratado desde un punto de vista técnico ni en profundidad, dados nuestros conocimientos.

Para analizar el funcionamiento de cada motor hemos elaborado la siguiente ecuación: "El Camino de Santiago a pie. Albergues, restaurantes y servicios médicos". Esta ecuación se ha tenido que modificar según los instrumentos de búsqueda, teniendo que reducirla en la mayoría de los casos a la frase "Camino de Santiago", para obtener suficientes enlaces.

ASPECTOS A EVALUAR – A	METABUSCADORES - 1
	Metacrawler
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>SI</p> <p>SI</p> <p>SI</p> <p>SI</p> <p>NO</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios <p>• Información general: Nació en <u>1994</u> de la mano E.Selberg y Oren Etzioni, ambos pertenecientes a la Universidad de Washington. Se unió a InfoSpace Network en el 2000.</p> <ul style="list-style-type: none"> • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <p>-Páginas telefónicas amarillas</p> <p>-Paginas telefónicas blancas</p> <p>-Servicios de tipo comercial</p> <p>-Entretenimiento y ocio</p> <p>-Breve directorio de temas más buscados</p> <p>-Admite personalizar el método de trabajo de forma permanente en la opción "Customize".</p> <p>-Puede ser tu herramienta de búsqueda por defecto.</p> <p>-MiniCrawler: da acceso constante a las principales webs con solo efectuar la consulta en su diminuta ventana.</p> <p>-MetaSpy: detecta lo que están buscando en MetaCrawler.</p> <p>-Con "Low Bandwidth Search" puedes agilizar tu búsqueda suponiendo que uses un browser lento y viejo.</p> <p>-<u>Las herramientas con las que trabaja son:</u> Alta Vista, DirectHit, FindWhat, Kanoodle, LookSmart, MetaCatalog, OpenDirectory, Overture, Sprinks by About.</p> <p>-¿Cuándo conviene usar un metabuscador?:</p> <ul style="list-style-type: none"> • al principio de una consulta, para tener una idea general • cuando se busca sobre una materia poco común • cuando no se quieren formular búsquedas complejas • cuando se quieran recuperar tantos documentos como sea posible con una sola pregunta <p>SI</p> <p>SI</p> <p>-</p> <p>SI</p>

OBSERVACIONES

Dado que la ecuación de búsqueda propuesta originalmente a penas si nos proporciona resultados, introducimos "Camino de Santiago" que es la que mejores resultados nos ha ofrecido.

ASPECTOS A EVALUAR – B	METABUSCADORES - 1 Metacrawler
<p>1. INTERFASE <i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados • Valoración • Menús claros • Ampliación de información 	<p>SI NO presenta frames NO SI. Se puede escoger la forma de verlos: por relevancia, por sitio, por fuente. NO SI SI</p>
<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis • Página de ayuda • Ayuda de calidad 	<p>SI -Cuatro casillas donde introducir: Todas las palabras (all) Alguna palabra (any) La frase exacta (phrase exacta) Selección por tipo de archivo -Se puede limitar la búsqueda por dominio, escoger los buscadores que se deseen de la lista que presenta, a un tiempo determinado de respuesta, a un número determinado de resultados/página NO -</p>
<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Sólo hemos obtenido 21 enlaces en total. 3 / 10 = 30% 3 / 21 = 14%; 3 / 100 = 3%</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>—</p>
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>InfoSpace lo gestiona. Metacrawler recibe la consulta del usuario en HTTP, y después la lanza (también via HTTP) y la transforma en la sintaxis específica de cada uno de los motores que utiliza. Espera la respuesta de los distintos motores en un tiempo determinado, borra duplicados, los lista, y los muestra al usuario. El código que permite hacer lo dicho es C++. La presentación de los resultados se hace con la combinación de PHP, mod_perl y Jserv. Como servidor de imágenes se usa "Zeus".</p>

ASPECTOS A EVALUAR – A	METABUSCADORES - 2	
	Dogpile	
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	SI	SI
<p><i>Contenido</i></p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general: Surgió en 1996. Es propiedad de InfoSpace. • Sistematización • Información de enlaces • Rutas • Palabras clave 	SI	-Páginas telefónicas amarillas -Páginas telefónicas blancas -Servicios de tipo comercial -Entretenimiento y ocio <u>-Trabaja con las siguientes herramientas:</u> Búsqueda en la web: About, Ah-ha, Ask Jeeves, Dogpile Web Catalog, ePilot, FindWhat, Kanoodle, LookSmart, Open Directory, Overture, SearchHippo, Sprinks an Yahoo!. Imágenes: Ditto Audio/MP3: Astraweb, Audio Galaxy an MP3Board Archivos: Yaga Noticias: Dogpile Web Catalog Multimedia: VastVideo SI SI – SI

OBSERVACIONES

Dado que la ecuación de búsqueda propuesta originalmente a penas si nos proporciona resultados, introducimos: camino+de+santiago que es la que mejores resultados ha ofrecido.

ASPECTOS A EVALUAR – B	METABUSCADORES - 2
	Dogpile
<p>1. INTERFASE <i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados <ul style="list-style-type: none"> • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>NO</p> <p>SI</p> <p>SI. En cada página presenta los resultados de los cuatro primeros buscadores que analiza. Si no hay, por lo menos, 10 documentos, automáticamente presentará los de los 4 buscadores siguientes, y así sucesivamente.</p> <p>NO</p> <p>SI</p> <p>NO</p>
<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis <ul style="list-style-type: none"> • Página de ayuda • Ayuda de calidad 	<p>SI, pero no presenta ventana a parte</p> <p>-AND o +, NOT o -, OR : Dogpile trabajará solamente con aquellos buscadores compatibles con estos operadores.</p> <p>-Comillas</p> <p>SI, MUY BUENA. "Arfie" es el perrito que simboliza la ayuda</p> <p>SI</p>
<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 100 primeros resultados 	<p>Se han obtenido 462 resultados.</p> <p>Hay que destacar que de los 462 resultados totales, 421 corresponden a Dogpile web catalog.</p> <p>De los 100 primeros resultan pertinentes 53, o sea, el 53%.</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	—
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>Dogpile pertenece a la compañía InfoSpace, Inc desde el año 2000, y desde ella es gestionado: ofreciendo servicios y ayuda en general.</p>

ASPECTOS A EVALUAR – A	ÍNDICES TEMÁTICOS - 1
	Yahoo
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>NO, hay demasiado recargamiento</p> <p>NO. Además la página cambia constantemente.</p> <p>SI</p> <p>SI</p> <p>NO</p>
<p><i>Contenido</i></p> <ul style="list-style-type: none"> • Autoría • Servicios <ul style="list-style-type: none"> • Información general: Nace en <u>1994</u> en la Universidad de Stanford, de la mano de David Filo y Jerry Yang. <ul style="list-style-type: none"> • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <p>-Páginas telefónicas amarillas</p> <p>-Servicios de tipo comercial</p> <p>-Entretenimiento y ocio</p> <p>-Yahooligans!: versión infantil de Yahoo!</p> <p>-Links a versiones específicas del buscador por países y regionales.</p> <p>-Personalización de la página principal de Yahoo! En "My", en función de las necesidades de cada uno.</p> <p>-Existe la posibilidad de crear una búsqueda y mantenerla para que el sistema periódicamente haga la ejecute.</p> <p>-Se puede usar "Usenet" para acceder directamente a la base de datos de un grupo de noticias mantenida por Deja</p> <p>- Lo que empezó siendo una mera recopilación de las páginas favoritas de sus creadores, se convirtió en uno de los buscadores más populares.</p> <p>-Su diseño y contenido corresponde más al de un Portal que al de un típico Índice temático .</p> <p>-Es el más adecuado cuando se necesita información general de un tema, o también, cuando no se está muy seguro de lo que se quiere.</p> <p>-Su Directorio aparece distribuído en 14 categorías temáticas.</p> <p>SI</p> <p>SI</p> <p>SI</p> <p>SI</p>

OBSERVACIONES

La ruta seguida a través de este índice para localizar documentos sobre "Camino de Santiago" ha sido: España/Galicia/A coruña/Ciudades/Santiago de compostela/Comunidad/Religión/El camino de santiago

ASPECTOS A EVALUAR – B	ÍNDICES TEMÁTICOS - 1
	Yahoo
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda <ul style="list-style-type: none"> • Resultados <ul style="list-style-type: none"> • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>NO tiene frames</p> <p>SI. Muy completa.</p> <p>-En "howto.yahoo.com" hay un amplio tutorial que incluye aspectos importantes a tener en cuenta, glosario de internet, ejemplos de búsquedas, etc.</p> <p>-Además, junto a las categorías puede aparecer "NEW": significa que hay en ella elementos nuevos.</p> <p>-"Most popular sites": orientan sobre aquellos sitios más solicitados.</p> <p>-unas "gafas" indican que una página es especialmente buena</p> <p>Los resultados que aparecerán primero serán los de la propia base de Yahoo!, y al final los encontrados por Google.</p> <p>Yahoo!, no es un buscador si no un Índice, por tanto, no indiza a texto completo los resultados de búsqueda, sino que muestra únicamente la categoría, título, y una breve descripción de los mismos.</p> <p>NO</p> <p>SI</p> <p>SI. Bajo los resultados hay un menú donde se puede ampliar la consulta: "temas relacionados" "Eventos en internet" (también relacionados)</p>
<p style="text-align: center;">2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis y refinamiento <ul style="list-style-type: none"> • Página de ayuda • Ayuda de calidad 	<p>SI</p> <p>-AND, OR, NOT, comillas, * con las 3 primeras letras</p> <p>-Palabras clave</p> <p>-No tiene en cuenta May o Min.</p> <p>-Limitar la búsqueda a título, URL, un período temporal determinado, en la web o en el directorio</p> <p>-Se puede cambiar el número de resultados por página.</p> <p>-Es posible evitar que los sitios comerciales aparezcan en los resultados de búsqueda.</p> <p>SI</p> <p>SI</p>

<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none">• 1ª Pantalla• 100 primeros resultados	24 resultados; 22 pertinentes = 91%
<p>4. RECALL</p> <ul style="list-style-type: none">• % de precisión de los 100	22/53 = 41,5%
<p>5. ARQUITECTURA Y DISEÑO:</p>	<p>En lugar de dejar que su base de datos sea alimentada por la información vertida por robots o programas spider, Yahoo! tiene más de 150 editores, que revisan y deciden qué información formará parte de esa base de datos. Cuando se hace una consulta en Yahoo!, la búsqueda se hace, de forma complementaria, tanto en ese Directorio como en la gigantesca base de datos de Google, poseedora de más de un billón de páginas web.</p>

ASPECTOS A EVALUAR – A	ÍNDICES TEMÁTICOS – 2
	L.I.I.
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>SI SI NO HAY SI NO</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios <p>• Información general: Nace en <u>1990</u>.</p> <ul style="list-style-type: none"> • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI -Enlaces de actualidad: con una suscripción puedes recibir semanalmente la últimas novedades. -Sugerencias. -Co-Branding: es un servicio a través del cual, todas las bibliotecas que son miembros de la Biblioteca de California, pueden ser provistas de una versión de LII adaptada a sus características particulares. -El personal de LII, cada año, ofrece la posibilidad de recibir un curso de indización a los profesionales de California. Es un directorio por orden alfabético. Tiene más de 9.000 recursos de Internet seleccionados y evaluados por expertos. SI. SI NO SI</p>

OBSERVACIONES

La ruta seguida a través de este índice para localizar documentos sobre "Camino de Santiago" ha sido: Travel / Outdoor Recreation / Hiking/

ASPECTOS A EVALUAR – B	ÍNDICES TEMÁTICOS - 2
	L.I.I.
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>No tiene frames.</p> <p>NO</p> <p>NO</p> <p>SI</p> <p>SI</p> <p>SI. Da la opción de usar, en caso necesario, otras herramientas de búsqueda: Buscadores: Google, AllTheWeb, AltaVista, HotBot Metabuscaros: MetaCrawler, Ixquick Directorios temáticos: LII, Yahoo!</p>
<p style="text-align: center;">2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis <ul style="list-style-type: none"> • Página de ayuda • Ayuda de calidad 	<p>SI</p> <p>AND, OR, NOT, *</p> <p>Búsqueda por: tema, título, autor, URL, categorías, palabras clave.</p> <p>Hay que tener cuidado con el tema, ya que se guían por el Índice de materias de la Library of Congress.</p> <p>SI</p> <p>SI</p>
<p style="text-align: center;">3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Ningún resultado válido.</p> <p>Ningún resultado válido.</p>
<p style="text-align: center;">4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p style="text-align: center;">—</p>
<p style="text-align: center;">5. ARQUITECTURA Y DISEÑO</p>	

ASPECTOS A EVALUAR – A	BUSCADORES - 1 TEOMA
<p style="text-align: center;">1. INTERFASE</p> <p>Aspectos formales y gráficos</p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<ul style="list-style-type: none"> SI SI NO hay SI SI
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general: Se fundó en el año <u>2000</u> en New Jersey de la mano de un grupo de científicos de la Universidad de Rutgers. Su antiguo nombre era Direct Hit. • Sistematización • Información de enlaces • Rutas • Palabras clave 	<ul style="list-style-type: none"> SI NO HAY -Para determinar la calidad de una página TEOMA utiliza "Subject-Specific Popularity": se basa en el número de páginas que la referencian. Utiliza, para ello, 3 técnicas exclusiva:Refinamiento-Resultados-Recursos. SI NO - SI

OBSERVACIONES

Hemos modificado la ecuación inicial de búsqueda para obtener resultados suficientemente pertinentes. Se ha introducido "Camino de Santiago".

ASPECTOS A EVALUAR – B	BUSCADORES - 1
	Teoma
<p>1. INTERFASE <i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados • Valoración • Menús claros • Ampliación de información 	<p>SI No tiene frames. NO NO SI. Existe la opción "Results" en donde los resultados aparecen en orden a la importancia que el "Subject-Specific Popularity" les ha dado SI SI. -En la opción "Resources" se puede acceder a una colección de links que considera importantes para nuestro tema, y que ha sido confeccionada por expertos de todo el mundo. -Además, hacen comentarios de cada uno de los recursos, y que pueden ser muy útiles para el usuario.</p>
<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis • Página de ayuda • Ayuda de calidad 	<p>NO -Admite comillas o guiones sin espacio para especificar una frase o también se puede usar la ventana de búsqueda activando "Find this phrase". -Por defecto, omite en las búsquedas palabras muy comunes. Excepto que se incluya un "+" o un "-" delante de cada palabra que se quiera incluir o excluir. -Cuanto más específico se es, mejor trabaja. -Conviene usar distintas terminaciones de una misma raíz. Ya que los resultados pueden variar considerablemente. -Se puede limitar la búsqueda en la opción "Refine": aquí, TEOMA presenta los sitios organizados en clusters de elementos del mismo tema. NO -</p>
<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Se han obtenido 25.000 enlaces. 3 / 10 = 30% 20 / 100 = 20%</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>20/53 = 37,7%</p>
<p>5. ARQUITECTURA Y DISEÑO</p>	

ASPECTOS A EVALUAR – A	BUSCADORES - 2
	Hotbot
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>SI</p> <p>SI</p> <p>NO HAY</p> <p>SI</p> <p>SI</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios: algunos gratis otros no. <p>• Información general: Apareció en <u>1996</u>, y actualmente es propiedad de Terra Lycos.</p> <ul style="list-style-type: none"> • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <p>-Canales</p> <p>-Grupos de discusión</p> <p>-Páginas amarillas, Páginas blancas</p> <p>-Correo electrónico</p> <p>-Compras</p> <p>-Comunidades</p> <p>-Mensajes a móviles</p> <p>-Traductor</p> <p>-Software</p> <p>-La opción "Personalize these settings" te permite quitar o poner según las necesidades.</p> <p>-Nació como herramienta de búsqueda principalmente. Por ello, los elementos típicos de los portales, aparecen relegados a un segundo plano.</p> <p>-Es muy popular debido a su fácil manejo, y a las grandes posibilidades que ofrece a la hora de hacer las búsquedas. Y en este sentido rivaliza con AltaVista, aunque tiene mucho menos tránsito que sus rivales. Es, por ello, una de las herramientas de búsqueda preferidas por los profesionales.</p> <p>-Tiene un pequeño directorio: Con 14 categorías, ha sido creado y mantenido por miles de editores voluntarios, coordinados por el Open Directory Project.</p> <p>SI</p> <p>SI</p> <p>–</p> <p>SI</p>

ASPECTOS A EVALUAR – B	BUSCADORES - 2
	Hotbot
<p>1. INTERFASE <i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda <p>• Resultados</p> <ul style="list-style-type: none"> • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>Sí tiene frames.</p> <p>-SI. La ayuda de Hotbot está organizada en preguntas más frecuentes.</p> <p>-Usando las "Meta Words" se facilita la búsqueda. Se ofrece una lista de ellas en la "Help".</p> <p>-SI. Cada resultado está encabezado por lo que Hotbot llama ""Full descriptions": menciona autor, título, 3 ó 4 líneas de comentarios, autor, URL, última modificación.</p> <p>-Se puede refinar la búsqueda dentro de los resultados obtenidos.</p> <p>NO</p> <p>SI</p> <p>-SI. Con la opción "This site only" puedes obtener más páginas, que es lo que viene por defecto, de cada sitio web.</p> <p>-Usando "Top 10 sites for Jeep grand cherokee" se puede acceder a los sitios más solicitados de acuerdo con tu consulta. Este servicio viene de TEOMA, que como vimos, ofrecía los resultados basándose en el número de veces que éstos eran referenciados.</p> <p>-Ofrece la posibilidad de utilizar "LYCOS" si no se ha podido resolver la consulta.</p>
<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis <p>• Página de ayuda</p> <p>• Ayuda de calidad</p>	<p>SI</p> <p>-AND, OR, NOT, frase exacta</p> <p>-Admite case-sensitive, Asterisco, comillas, interrogación, paréntesis.</p> <p>-Se puede limitar la búsqueda por idiomas, título, URL, nombres de persona, por fecha, por tipo de archivo (image, MP3, Video, JavaScript), por número de páginas por sitio.</p> <p>-Se puede limitar el número de resultados/página.</p> <p>SI</p> <p>SI</p>
<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Se han obtenido aprox. 25.000 enlaces.</p> <p>$3 / 10 = 30\%$</p> <p>$20 / 100 = 20 \%$</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>$20/53 = 37,7\%$</p>
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>Inktomi creó el software que utiliza HotBot para peinar la web y conformar posteriormente la base de datos que usa, y que se llama Slurp.</p> <p>Además de HotBot, AOL Search, y MSN también son llevados por Inktomi, pero HotBot es de los pocos en ofrecer su base de datos GEN3 que es mucho mayor.</p>

ASPECTOS A EVALUAR – A	<i>BUSCADORES - 3</i>	
	All the web	
<p>1. INTERFASE <i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	SI	SI
<p><i>Contenido</i></p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general • Sistematización • Información de enlaces • Rutas • Palabras clave 	SI	NO
	-	SI
	SI	SI
	SI	SI

OBSERVACIONES:

La ecuación de búsqueda que se ha propuesto es:

Frase exacta: Camino de santiago a pie. Y especificamos que debería incluir: albergues; restaurantes; servicios médicos.

ASPECTOS A EVALUAR – B	BUSCADORES - 3
	All the web
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>SI EVITA SU USO</p> <p>SI. Existe la opción "Search Tip" muy útil y original: te alerta de aspectos importantes de la búsqueda concreta que haces.</p> <p>SI.</p> <p>-Cada resultado aparece con el título, una breve explicación, y la URL de la página correspondiente. Se puede prescindir de todo ello, excepto del título en la opción "Title Only".</p> <p>-Por defecto, aparecen 10 resultados por página.</p> <p>-Los resultados se pueden guardar, retocar, borrar, re-intentar la búsqueda</p> <p>NO</p> <p>SI. Los resultados aparecen agrupados por categorías para facilitar la navegación. Apretando en cada una de ellas, aparecen todos los resultados que tienen que ver con ella.</p> <p>SI.</p> <p>Apretando "More hits from" se puede realizar una nueva búsqueda con la misma ecuación pero solamente dentro del sitio en el que se está.</p>
<p style="text-align: center;">2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis • Página de ayuda • Ayuda de calidad 	<p>SI</p> <p>→Hay 1 ventana de búsqueda donde se despliegan 3 opciones:</p> <ul style="list-style-type: none"> - todas las palabras (equivaldría a un AND) - algunas de las palabras (OR) - la frase exacta <p>→También da la posibilidad de incluir filtros para ciertas palabras. Hay 3 ventanas, cada una de las cuales se puede repetir tantas veces como se desee:</p> <ul style="list-style-type: none"> - la 1ª: debería incluir - la 2ª: debe incluir - la 3ª: no debe incluir <p>→En Alltheweb se da la opción de buscar en 6 diferentes medios: web, noticias, fotografía, videos, audio, archivos FTP.</p> <p>→Da la opción de aplicar cada una de ellas en distintas partes: texto, URL, título, etc.</p> <p>→Se puede limitar la búsqueda por :</p> <p>Lengua, dominio, dirección IP, periodo cronológico, tamaño del documento.</p> <p>SI</p> <p>SI</p>

<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>149 enlaces $9/10 = 90\%$ $33/100 = 33\%$</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>$33/53 = 62,3\%$</p>
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>Fast se encarga de gestionar la información para ALLTHEWEB y otros muchos: por ejemplo se encargará de indizar todo un sitio web o solamente aquellas páginas que el cliente quiera que se vean. El software patentado por FAST es capaz de tratar con gran cantidad de información desde una sola CPU. Agrupa los resultados entorno a 2 ejes: volumen de datos y número de usuarios simultáneos. Desarrolla un filtrado a tiempo real y compresión de productos y servicios basado en su plataforma tecnológica modular, que tiene 3 niveles:</p> <p>1º → reunión y procesamiento de datos: indexación y procesamiento en bases de datos propietarias utilizando algoritmos de gran velocidad.</p> <p>2º → investigación y filtrado a tiempo real: arquitectura distribuida</p> <p>3º → Presentación y distribución: los datos le son ofrecidos al usuario en el formato que desee, aunque se usan formatos estándar.</p>

ASPECTOS A EVALUAR – A	BUSCADORES - 4 Altavista
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>Demasiado recargamiento. Lo opuesto a p.e. Google u Alltheweb. Sin embargo, existe la posibilidad de prescindir de toda información que no sea la puramente de búsqueda de información, a través de la opción "Raging search".</p> <p>Mucha publicidad y todo aquello que es característico de un portal.</p> <p>No excesiva. Está constantemente cambiando su página principal. Lo que hoy te han ofrecido, mañana puede no estar ya.</p> <p>SI</p> <p>SI</p> <p>NO</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general: Nace en <u>1995</u> de la mano del Digital Equipment Corp., en California. Mayoritariamente es propiedad de CMGI y en menor parte de Compacq. • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <ul style="list-style-type: none"> -Filtro familiar -Ocio, compras -Páginas amarillas -Información de tipo comercial -Traductor Babel fish, usando el link "Translate link" -Posibilidad de hacer de Alta vista tu página por defecto -"My Alta Vista" te permite personalizar tu versión. -Es el buscador favorito de los profesionales por su gran capacidad para hacer "a medida" las opciones de búsqueda, así como por su gran base de datos. -Empezó por ser un motor de búsqueda simplemente, pero con el tiempo se ha ido convirtiendo en un gran portal con todas las características propias del mismo. -Se puede utilizar el directorio cuando no se sabe exactamente qué queremos, pero no sin hacer uso de operadores booleanos. <p>SI</p> <p>SI</p> <p>SI</p>

OBSERVACIONES: La ecuación de búsqueda es "camino de santiago a pie" AND (albergues OR restaurantes OR servicios médicos)

ASPECTOS A EVALUAR – B	BUSCADORES - 4
	Altavista
<p>1. INTERFASE <i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados <ul style="list-style-type: none"> • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>SI tiene frames.</p> <p>SI</p> <p>-SI. También incluye el título del documento, una breve descripción en la que incluye al autor del mismo, la URL, y todos los términos de búsqueda aparecen destacados.</p> <p>-Alta Vista te da la opción de cambiar la forma de ver los resultados en la ventana "Results Opciones": descripción, URL, última modificación, tamaño página web, traducción, etc..</p> <p>NO</p> <p>SI</p> <p>SI. Hace clusters de los resultados y por cada sitio, por defecto, solamente presenta una página. Si se quiere ampliar la lista de resultados por sitio se puede usar "More pages from this site".</p> <p>Se puede intentar una nueva búsqueda, modificando algún término, dentro de los primeras páginas, si no se ha dado exactamente con lo que se quiere, pero los primeros resultados se ve que están en la onda: "Search within these results".</p> <p>Se puede usar la opción "Related searches" para ampliar la búsqueda a páginas relacionadas con el tema.</p>
<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis <ul style="list-style-type: none"> • Página de ayuda <ul style="list-style-type: none"> • Ayuda de calidad 	<p>SI</p> <p>AND, OR, AND NOT, NEAR, paréntesis, comillas, admite "Case sensitive", *.</p> <p>Se puede limitar la búsqueda a una sola lengua, a un cierto número de resultados/pág., por zonas geográficas, sitios web, por dominios</p> <p>Se puede especificar dónde buscar: en páginas Multimedia, Comerciales, Noticias, Ocio, Negocios, Educación, Política.</p> <p>SI. En "Ata vista advanced search quick reference" se ofrece un resumen de cómo manejarse adecuadamente.</p> <p>SI</p>

<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none">• 1ª Pantalla• 100 primeros resultados	<p>Hemos obtenido 9 enlaces. Pertinentes: 6 $6/10 = 60\%$ $6/53 = 11,3 \%$</p>
<p>4. RECALL</p> <ul style="list-style-type: none">• % de precisión de los 100	
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>-La base de datos ha sido creada por Scooter: programa de software que pertenece a Alta Vista y que constantemente recorre la web para dar de alta páginas nuevas.</p> <p>-Cada día incorpora a Alta Vista aproximadamente 10 millones de páginas nuevas. Después de lo cual, un software muy potente y especial indexa cada palabra. Al trabajo de Scooter se suma la aportación de los creadores de sitios web.</p> <p>-Así como es muy bueno incorporando páginas nuevas, carece de un buen mecanismo cuando se trata de eliminar las páginas muertas.</p>

ASPECTOS A EVALUAR – A	BUSCADORES - 5 Google
<p style="text-align: center;">1. INTERFASE</p> <p>Aspectos formales y gráficos</p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>SI. Muy bueno. No hay elementos que distraigan del propósito exclusivo de la búsqueda de información. Opciones de búsqueda más sofisticadas aparecen relegadas a un segundo plano. No hay elementos publicitarios por defecto.</p> <p>SI</p> <p>SI</p> <p>SI</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general: Creado en <u>1998</u> por los estudiantes Larry Page y S.Brin en Stanford. • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <p>-SI. Se ofrece la posibilidad de diseñar la página de búsqueda al gusto del usuario en "Preferencias".</p> <p>-También Google puede aparecer como primera herramienta de búsqueda en nuestro ordenador.</p> <p>-Su éxito viene por la claridad de la interface, la rapidez de respuesta y la libertad que ofrece a la hora de hacer las búsquedas.</p> <p>-Su base de datos es la más grande de todas, con más de 1.2 billones de páginas.</p> <p>Su directorio está organizado, (no se accede directamente desde la página principal), en 16 grandes categorías, que se alimentan del Open Directory Project.</p> <p>No numera los resultados</p> <p>SI</p> <p>SI.</p> <p>SI</p>

OBSERVACIONES: La ecuación de búsqueda que se ha propuesto es:

Frase exacta: camino de santiago a pie

Algunas palabras: restaurantes albergues servicios médicos

ASPECTOS A EVALUAR – B	BUSCADORES - 5 Google
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda <ul style="list-style-type: none"> • Resultados <ul style="list-style-type: none"> • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>SI las evita.</p> <p>-SI, muy buena. Hay un elemento de ayuda muy importante, "Cached": si al conectar con una página, surgen problemas, a través de esta opción se puede acceder a una copia de la última actualización de la página.</p> <p>-Son muy útiles "All about Google" o "Search tips" para obtener toda la ayuda necesaria.</p> <p>-SI. Proviene básicamente de su gran base de datos, así como del directorio creado por el "Open directory Project" que se alimentan de las aportaciones que hacen personas y no máquinas.</p> <p>-Con la opción "I'm feeling lucky" en lugar de presentar toda la lista de resultados, muestra el que cree más pertinente.</p> <p>-Aparecen muy bien marcados y también los términos de búsqueda.</p> <p>-Es el único motor que indexa a texto completo archivos PDF y los incluye en los resultados de búsqueda.</p> <p>-Como los demás, muestra el número de elementos encontrados, pero además, el tiempo que tardó en hacerlo.</p> <p>NO</p> <p>SI</p> <p>-SI. Da la oportunidad, al igual que muchos otros, de ampliar el número de página por sitio utilizando "More results".</p> <p>-Incluye la opción "Similar pages" para buscar más información en otras páginas que considere están relacionadas con el tema.</p> <p>-Un sistema más consistente para encontrar páginas relacionadas es usar "category" para conectarse con el directorio a través de los links de las categorías.</p>

<p>2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis <ul style="list-style-type: none"> • Página de ayuda • Ayuda de calidad 	<p>SI</p> <p>-Google no te da tantas posibilidades de preparar una ecuación de búsqueda como p.e Alta Vista, HotBot o incluso Northern Light: No se puede usar otro operador booleano que el “OR”, tampoco admite “case-sensitive” o el “*”, las frases hechas dan problemas. Pero sin embargo, es capaz, con 2 ó 3 palabras clave, de ofrecerte lo que buscas.</p> <p>-El procedimiento no es metiendo los operadores, ya que los incluye por defecto en las distintas casillas.</p> <p>-Si</p> <p>-Admite limitar la búsqueda a través del título, URL, dominio, links y lengua.</p> <p>-También permite limitar el número de resultados/pág.</p> <p>-Permite la búsqueda en herramientas de búsqueda más específicas: Apple Macintosh, BSD, Unix, Linux, universidades, etc..</p> <p>SI</p> <p>SI</p>
<p>3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Se han obtenido 65 resultados. Pertinentes: 33</p> <p>$9/10 = 90\%$</p> <p>$33/65 = 50\%$</p>
<p>4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>$33/53 = 62,3\%$</p>
<p>5. ARQUITECTURA Y DISEÑO</p>	<p>-Google trabaja con una combinación única de software y hardware muy sofisticado. La gran velocidad que alcanza en parte se debe al algoritmo que usa pero también a la enorme red de pequeños ordenadores conectados a él.</p> <p>El corazón de su software es PagRank&Trade, un sistema para listar las páginas web creado por sus fundadores.</p> <p>-Google alcanza gran relevancia en sus resultados de búsqueda: emplea sofisticadas técnicas de análisis de links y de significado de los datos, que determinarán el orden de aparición de los mismos. Por tanto, la relevancia viene dada por el número de links que una página recibe y ese será el factor de selección para Google: para él un link desde la página A a la página B, será un voto para B. Pero Google también analiza de dónde proviene ese link, de manera que los que considera más importantes los pondera más.</p> <p>-Tanto Yahoo como Netscape Search, utilizan su tecnología. De manera que cuando se trabaja con estos sitios, los resultados, por lo menos en parte, pueden venir de la base de datos de Google.</p>

ASPECTOS A EVALUAR – A	BUSCADORES - 6 Northern light
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos formales y gráficos</i></p> <ul style="list-style-type: none"> • Consistenc., claridad y legibilidad • Estabilidad • Imágenes • Enlaces visitados • Enlaces numerados 	<p>SI</p> <p>SI</p> <p>SI. Hay pocas.</p> <p>SI</p> <p>SI</p>
<p>Contenido</p> <ul style="list-style-type: none"> • Autoría • Servicios • Información general: Creado en <u>1997</u> por un grupo de bibliotecarios, documentalistas e informáticos de Massachusetts. • Sistematización • Información de enlaces • Rutas • Palabras clave 	<p>SI</p> <p>Su servicio de alerta es de los mejores: informa de lo último sobre un tema y te permite guardarlo.</p> <p>Servicios para empresas.</p> <p>Noticias de actualidad.</p> <p>Noticias financieras.</p> <p>Artículos a texto completo pagando.</p> <p>-No se trata de un buscador de páginas web, sino de un banco documental que recoge documentos en la red, los indiza, clasifica y pone a disposición del usuario en forma de pago.</p> <p>-Bastante orientado a empresas, negocios, comercio.</p> <p>-Documentos de tipo técnico.</p> <p>-Special Collection: la mejor información de cada tema, a texto completo.</p> <p>SI</p> <p>Cada documento aparece catalogado.</p> <p>–</p> <p>SI</p>

ASPECTOS A EVALUAR – B	BUSCADORES - 6 Northern light
<p style="text-align: center;">1. INTERFASE</p> <p><i>Aspectos técnicos. Navegación</i></p> <ul style="list-style-type: none"> • Software • Frames • Ayuda • Resultados • Valoración • Menús claros • Ampliación de información 	<p>SI</p> <p>Complementaria a la búsqueda.</p> <p>SI</p> <p>-SI. Existe la opción "Custom Search folders": carpetas que organizan los resultados por temas, fuente, tipo de documento, lenguaje.</p> <p>-En los resultados se incluirán tanto los que vienen de la web como los de la colección especial: los de la web ofrecerán una breve descripción, fecha de actualización, tipo de sitio. Los especiales: tipo de documento, fecha de publicación y fuente.</p> <p>SI. En cada resultado aparece un % que indica su pertinencia.</p> <p>SI</p> <p>-SI. Tener en cuenta la Colección Especial y que pagándolos se pueden tener a texto completo.</p> <p>-Existe la opción "More results" para ver más páginas de un sitio concreto.</p>
<p style="text-align: center;">2. BÚSQUEDA AVANZADA</p> <ul style="list-style-type: none"> • Disponible • Sintaxis • Página de ayuda • Ayuda de calidad 	<p>SI</p> <p>-Operadores booleanos (AN, OR, NOT, +, -)</p> <p>-Entrecorillado, truncamientos, % para sustituir un solo carácter.</p> <p>-No ignora las palabras vacías, siempre que se entrecorillen.</p> <p>-No trabaja con "case-sensitive": es insensible a las May y Min.</p> <p>-Se puede limitar la búsqueda por temas, fuentes (colección especial o la web), tipos de documentos, fechas, URL, dominio, nombre de la publicación.</p> <p>-Se puede limitar el número de resultados/página.</p> <p>SI</p> <p>SI</p>
<p style="text-align: center;">3. PRECISIÓN / PERTINENCIA</p> <ul style="list-style-type: none"> • 1ª Pantalla • 100 primeros resultados 	<p>Ningún resultado pertinente.</p> <p>Ningún resultado pertinente.</p>
<p style="text-align: center;">4. RECALL</p> <ul style="list-style-type: none"> • % de precisión de los 100 	<p>---</p>
<p style="text-align: center;">5. ARQUITECTURA Y DISEÑO</p>	<p>Los "Custom Search Folders" han sido creados por Northern Light Software: han sido muy innovadores pero hay que tener en cuenta que no han sido compilados por personas.</p>

V. CONCLUSIONES

- Se observa que en aspectos como la sintaxis hay falta de homogeneidad y nula normalización, lo que crea bastante desorientación y confusión en el usuario a la hora de manejar estas herramientas.
- Las diferencias entre motores y directorios se van desdibujando, de manera que en los directorios temáticos se incorporan ventanas para hacer búsquedas por palabras clave, y en algunos de los motores hay la posibilidad de usar directorios temáticos.
- Aunque, salvando excepciones, la mayoría nacieron con propósito de ser meras herramientas de búsqueda, han sido empujados hacia posiciones más comerciales, pasando muchos de ellos a convertirse en auténticos portales.
- La presentación resulta excesivamente recargada y distrae al usuario.
- El carácter básicamente comercial, y la gran competencia que hay entre unos y otros, hace que cada empresa organice su producto de forma muy particular, lo que hace que, ante una misma búsqueda, se obtengan resultados absolutamente dispares, e incluso, incoherentes.
- Algunos motores destacan sobre otros, bien por su versatilidad a la hora de introducir los términos de búsqueda (como AtaVista), bien por la calidad y rapidez en la recuperación (Google).
- Dado el ingente número de páginas web que cada día se incorpora a la red, es urgente la especialización temática de los buscadores y una diferenciación clara entre aquellos de propósito general o más comerciales y aquellos de carácter más profesional o técnico.