Diseño de corpus y software para terminología

Introducción

- Estrecha relación entre la terminología y la informática y su influencia en la metodología.
- Procesamiento de la terminología está basada en corpus: parámetros para diseñar un corpus específico.
- Tecnologías para las tareas relacionadas con la terminología.

Índice

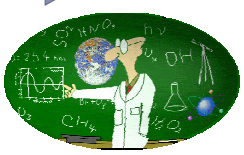
Diseño de corpus y software para terminología

Contenido– ¿Qué es terminología?

- **Terminología**

ciencia

Estudio científico de los conceptos y términos que hallan en los lenguajes de especialidad
(ISO 1087: 2000, 12)

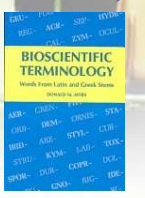


práctica

Directrices empleadas en terminografía que se refiere a su metodología

objeto

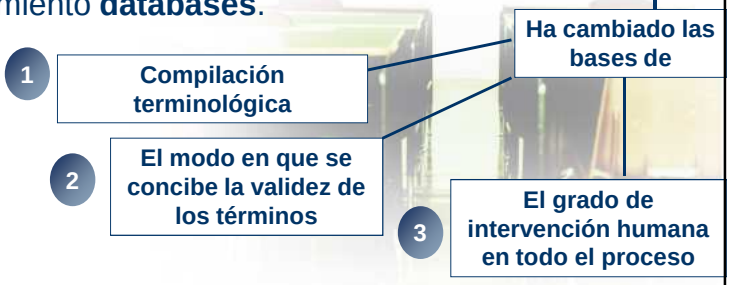
Conjunto de términos de una especialidad



Contenido- Informática y metodología en terminología

El efecto de la informática en la metodología terminológica:

- Uso de **corpus electrónicos** previamente registrado
- Explotación de **bases de datos** terminológicas y de conocimiento **databases**.



Diseño de corpus y software para terminología

Contenido- Informática y metodología en terminología

El efecto de la informática en la metodología terminológica:

- **RESULTADO**:
  - Acercamiento de la terminología y lexicografía

Terminology and lexicography are in theory two quite different disciplines, yet in practice they are headed towards a single methodology because of their common use of computer applications. (Sager, 1990)

  - La posibilidad de acceder a bancos de datos ha conllevado un cambio de foco en el trabajo terminológico.

Contenido- Informática y metodología en terminología

- Más información disponible para:
  - Traductores
  - Terminólogos

Ahora pueden estar más seguros/as de las decisiones terminológicas

- La posibilidad de acceder a grandes bancos de datos ha hecho que el trabajo terminológico sea:
  - Más sencillo
  - Más complejo

Terminology has gone from being an art to being a technique (Sager, 1990)

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

- 5 puntos básicos donde la informática representa un papel primordial para terminológos:

1. Selección de información previa al comienzo del trabajo
2. Creación de corpus y extracción de datos
3. Redacción de la entrada
4. Comprobación de la información en la entrada
5. Ordenación de las entradas terminológicas

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

1. Selección de la información

Los terminológos tienen acceso a verios tipos de bancos de datos con información previa general;

1.1 bancos documentales (publicaciones y diccionarios sobre el tema)

1.2 bancos textuales con corpus de textos técnicos;

1.3 bancos terminológicos con listados de términos sobre el tema e información lingüística e interlingüística de cada término.

2. Creación de corpus y extracción de datos

3. Redacción de la entrada

4. Comprobación de la información en la entrada

5. Ordenación de las entradas terminológicas

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

1. Selección de la documentación

2. Creación de corpus y extracción de datos

2.1 Los terminólogos pueden **seleccionar textos electr.** E incorporarlos en un banco de datos textual

2.2 Los textos en papel pueden convertirse en electrónicos con un **escáner y un OCR**

2.3 Los **textos** electrónicos pueden **analizarse** por programas semiautomáticos de extracción a fin de **detectar y extraer términos**

3. Redacción de la entrada

4. Comprobación de la información en la entrada

5. Ordenación de las entradas terminológicas

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

1. Selección de la documentación
2. Creación de corpus y extracción de datos
3. **Redacción de la entrada**

3.1 Elaboración de fichas terminológicas: los terminólogos usan los archivos para escribir fichas terminológicas transfiriendo la información procedente de los textos electrónicos de referencia (e.j. entrada, fuente, contexto, definición, etc.)

3.2 Una vez la ficha está completa, puede editarse total o parcialmente, o puede fusionarse con fichas de otras fuentes

3.3 Los terminólogos pueden controlar la información sobre referencias cruzadas and equivalencias.

4. Comprobación de la información en la entrada
5. Ordenación de las entradas terminológicas

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

1. Selección de la documentación
2. Creación de corpus y extracción de datos
3. Redacción de la entrada
4. **Comprobación de la información en la entrada**

Para comprobar y validar las entradas, los terminólogos revisan las bases de datos y transfieren cualquier información que falte a la ficha.

5. Ordenación de las entradas terminológicas

Diseño de corpus y software para terminología

Contenido – puntos básicos en terminografía

1. Selección de la documentación
2. Creación de corpus y extracción de datos
3. Redacción de la entrada
4. Comprobación de la información en la entrada
5. Ordenación de las entradas terminológicas
Para publicar, los ordenadores permiten al terminólogo presentar la información en varios formatos (papel, archivo electrónico, CD, etc.), a fin de incluir los datos que cada caso requiera.

Diseño de corpus y software para terminología

Contenido- corpus electrónicos y especializados

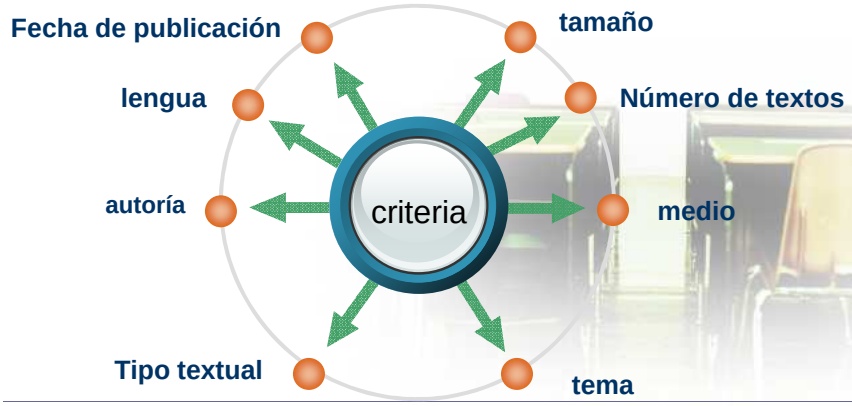
- **Una búsqueda por un término simple que no está en un diccionario implica:**
  - Seleccionar bilbiografía
  - Consultar a expertos en el ámbito

- **El acceso a bases de datos terminológicos and bases de datos de textos especializados ha modificado significativamente el método de trabajo**

- **En la actualidad, la terminología sistemática se base en la investigación sobre corpus representativos**

Contenido- corpus electrónicos y especializados

❖ Directrices para el diseño de corpus especializados:



Diseño de corpus y software para terminología

Contenido- corpus electrónicos y especializados

❖ Corpus:

- Colección de textos que compilados según criterios específicos

❖ Estos criterios se determinan según:

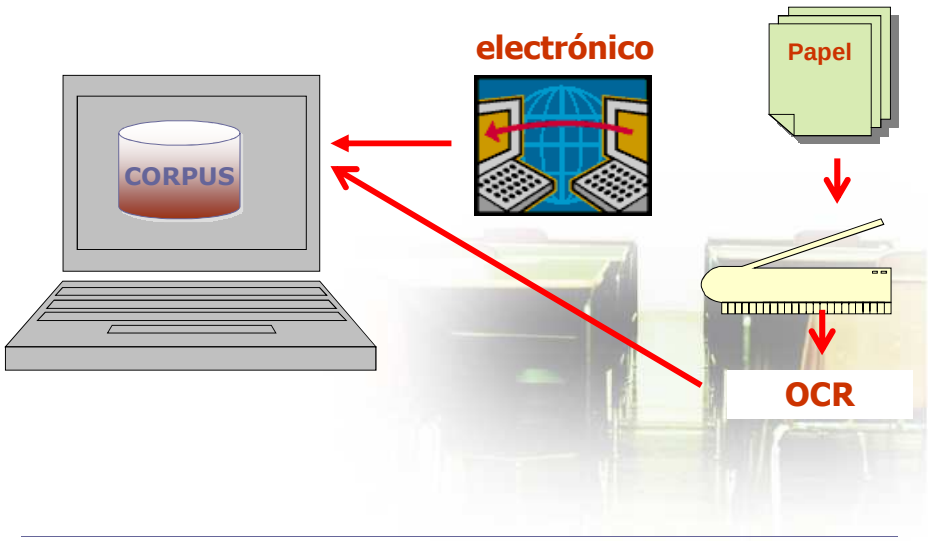
- Las necesidades del usuario
- Los propósitos del corpus

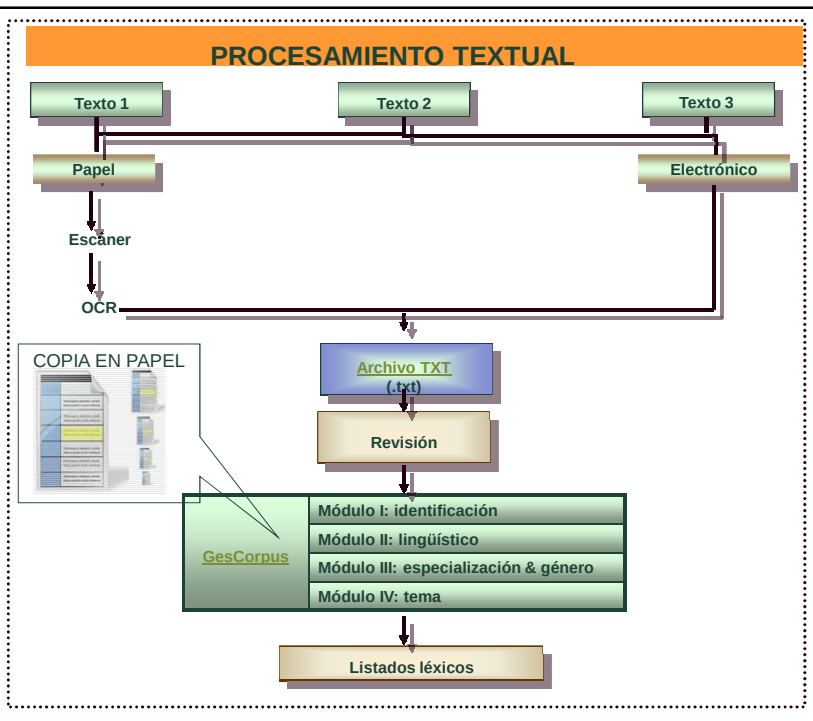
a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language (Sinclair, 1996)

(adoptada por European Advisory Group on Language Engineering Standard)

Diseño de corpus y software para terminología

Contenido- corpus electrónicos y especializados





Contenido– TAMAÑO DEL CORPUS

- Se hace alusión a 'gran' colección de textos (pero 'gran' resulta una noción vaga)
- No hay reglas que ayuden a determinar el tamaño ideal de un corpus.
- En su lugar, las decisiones se basan en aspectos como:
  - Las necesidades del proyecto
  - La disponibilidad de los datos
  - La cantidad de tiempo con la que contamos
- Es importante no asumir que cuanto más grande mejor
- La información útil pueden encontrarse en un corpus pequeño pero bien contruido diseñado para satisfacer las necesidades del usuario

Diseño de corpus y software para terminología

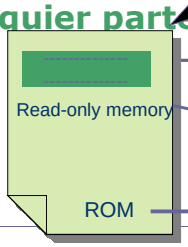
Contenido– Fragmentos de textos vs. textos completos

- El algunos corpus especializados, se limita el tamaño de los textos:
  - E.j. Lancaster-Oslo-Bergen corpus: fragmentos de sólo 2.000 palabras.
- En terminología, conceptos, términos, patrones y contextos pueden aparecer en cualquier parte del texto

Contenido– Fragmentos de textos vs. textos completos

- **El algunos corpus especializados, se limita el ta… …os:**
  - E.j. Lancas… …us: fragmentos de sólo 2.0…
- **En terminolo… …minos, patrones y co… …pueden aparecer en cualquier parte del texto**

Cuando se compila corpus especializados se emplean textos completos



- Concepto definido en un párrafo del principio.
- Un término completo puede emplearse al principio.
- Su abreviatura puede salir más tarde

Contenido– NÚMERO DE TEXTOS

- **Es también importante considerar:**
  - Cuántos textos diferentes se van a incluir en el corpus.
  - Cuántos de estos textos han sido escritos por autores diferentes.

E.j.

Corpus de 20.000 palabras

Compuesto de

Corpus varios autores
20 ≠ textos
20 ≠ autores

≠

Corpus de un autor
2 textos largos
Mismo autor

Diseño de corpus y software para terminología

Contenido– NÚMERO DE TEXTOS

- **Es también importante considerar:**
  - Cuántos textos diferentes se van a incluir en el corpus.
  - Cuántos de estos textos han sido escritos por autores diferentes.

Se obtiene una buena idea de qué términos se emplean habitualmente en el corpus de estudio

Expuesto a términos que un autor concreto prefiere

Compuesto de

Corpus varios autores

20 ≠ textos

20 ≠ autores

≠

Corpus de un autor

2 textos largos

Mismo autor

Diseño de corpus y software para terminología

Contenido - MEDIO

- **Se refiere a si el texto se ha preparado originalmente como:**
  - Texto escrito
  - Texto hablado (transcrito)

TRADUCTORES Y TERMINÓLOGOS

Contenido - TEMA

- **Los textos del corpus necesitan tratar sobre el ámbito de estudio**
  - Esto no siempre es tan fácil como parece
- **¿Por qué?**
  - Muchos temas especializados son multidisciplinares (bioquímica)
- **¿Dónde termina la bioquímica y empieza la química?**
  - Puede ser difícil saber dónde termina un ámbito especializado y empieza

Diseño de corpus y software para terminología

Contenido – Tipo textual

Diferentes en cuanto a estilo y vocabulario

TRADUCTOR:

- Ha de traducir un **artículo de investigación** sobre el **asma**:

- Sacará provecho de un corpus que contenga **otros artículos del mismo tema** (el asma)

- Estos textos contendrán ejemplos de **vocabulario y estilo apropiados al tipo textual**

Diseño de corpus y software para terminología

Contenido – Tipo textual

Diferentes en cuanto a estilo y vocabulario

TERMINÓLOGO:

- Ha de compilar un **glosario amplio** de términos empleados en el ámbito del **asma**:

- Para asegurar una cobertura lingüística y conceptual más completa, este profesional necesitará compilar un corpus que incluya una **variedad de tipos de textos** que versen sobre el tema del asma.

Diseño de corpus y software para terminología

Contenido - AUTORÍA

- **Para obtener un corpus con material especializado auténtico:**
  - El autor de cada texto debe ser un experto reconocido
  - Puede ser más fácil reconocer el autor de un texto impresothe author of a printed document, pero cuando usamos la web puede resultar más complejo.
  - En la web:
    - Una página personal puede ser menos fiable que
    - Un texto en un sitio web de una organización profesional reconocida.

Contenido - AUTORÍA

- **Para obtener un corpus con material especializado auténtico:**
  - El autor de cada texto debe ser un experto reconocido
  - Puede ser más fácil reconocer el autor de un texto impresothe au do usamos la wel
  - En la web:
    - Una página
    - Un texto en nal reconocida.

Formación profesional o académica apropiada para la disciplina sobre la que escribe

Reconocido por sus colegas por tener el nivel de pericia requerido para escribir sobre un tema concreto

Contenido - LENGUA

- **Usar material en lengua original proporciona al usuario ejemplos auténticos de usos típicos en un lenguaje de especialidad dado**

Diseño de corpus y software para terminología

Contenido– FECHA DE PUBLICACIÓN

- **Para el estado actual de un tema (plano lingüístico y conceptual)**

INCLUIR PRINCIPALMENTE TEXTOS ACTUALIZADOS

Diseño de corpus y software para terminología

Contenido– Corpus electrónicos y especializados

Los puntos principales hasta aquí…

1. Las directrices que nos ayudan a diseñar un corpus son: tamaño, nº de textos, medio, tema, tipo textual, autoría, lengua, fecha de publicación

2. Un corpus es una colección de textos electronicos compilados de acuerdo con unos criterios específicos

Contenido: Estación de trabajo terminológica y software

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

❖ Gestión de proyectos

- Herramienta que:
  - Garantice la coherencia del desarrollo del producto o servicio
  - Mejore la planificación, la comunicación, la coordinación y el trabajo en equipo
- Ejemplos:
  - Microsoft Project
  - Project Open (http://www.project-open.com/).

  (Puede consultarse una relación comparada de este tipo de software en http://en.wikipedia.org/wiki/Comparison_of_project_management_software)

Contenido: Estación de trabajo terminológica y software

Investigación

- Herramienta que:
  - Recupere, almacene y mantenga los textos del corpus
  - Debe permitir añadir, editar, borrar textos y documentarlos

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Investigación:

- un rastreador y recuperador automático de ficheros web
- un clasificador de documentos, según lengua, tema, grado de especialización
- un extractor de palabras clave
- un conversor a texto plano
- un sistema que permita editar documentos, clasificarlos e incluir información sobre su filiación
- una herramienta que mida la similitud entre documentos
- una base de datos factográfica
- una herramienta para crear estructura de conceptos

Contenido: Estación de trabajo terminológica y software

Explotación:

- un alineador de textos paralelos;
- un etiquetador;
- un extractor híbrido de terminología
- un programa de concordancias (monol. y paral.).

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Desarrollo:

- Un identificador automático de términos (la herramienta destaca en un txt nuevo las unidades léxicas ya recogidas en la bd);
- un sistema gestor de bases de datos terminológicas (adaptado a la combinatoria); y
- una herramienta para la preparación y publicación del trabajo en web o para su maquetación en papel.

❖H. bloque investigación:

Nombre	SO	Distribución	Lenguas
BootCaT front-end	Windows, Ubuntu, Debian, MacOSX, Linux	gratuita	26, entre las que incluye el inglés, español, francés, catalán, portugués, alemán…
	http://bootcat.sslmit.unibo.it/?section=frontend		
WebBootCaT	Integrado como un formulario dentro de la herramienta Sketch Engine	Comercial, por suscripción	42, entre las que incluye el inglés, español, francés, catalán, portugués, alemán…
	http://beta.sketchengine.co.uk/login/		
WeBoCa	Todos, aplicación en Java	gratuita	No se especifica
	http://code.google.com/p/weboca/		
WebAsCorpus.com	Aplicación en web	gratuita	40, entre las que incluye el inglés, español, francés, catalán, portugués, alemán…
	http://webascorpus.org/searchwac.html		
Webgetter	Windows. Utilidad de WordSmith Tools	comercial	todas
	http://www.lexically.net/wordsmith/version5/index.html		
TerminoWeb, v.2.0	Aplicación en web	gratuita, con suscripción	inglés y francés
	http://terminoweb.iit.nrc.ca/terminoWeb-v2_e.html		
Terminus	Integrado como un formulario dentro de la herramienta Terminus	comercial	
	http://melot.upf.edu/Terminus2009/index_es.html		

Diseño de corpus y software para terminología

Contenido: Estación de Trabajo terminológica

❖H. bloque investigación: extractores de palabras clave

- Principio: adquisición textual por Internet
- Antes de descargar conocer palabras frecuentes
- Por tanto, aplicación integrada en el rastreador
- Keyword Analysis Tool y AnalogX Keyword Extractor

Contenido: Estación de trabajo terminológica y software

H. bloque investigación: conversores a txt

- Los rastreadores los incorporan
- DoctoText (.doc, .xls, .rtf., .odf y xml) (http://silvercoders.com/en/products/doctotext)
- PDF to text: (http://www.pdftotext.net/)

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

H. bloque investigación: gestores de corpus

- No de análisis textual sino sistemas que editan documentos, los clasifican, registran información administrativa, de idioma, autoría, etc.
- En terminología:
  - Corpógrafo (Linguateca, Portugal)
  - Terminus (IULA)

Contenido: Estación de trabajo terminológica y software

❖ H. bloque investigación:

- Mapas conceptuales
  - **Terminus**
  - **IHMC CmapTools**, gratuito, se instala de manera local
  - **MapasConceptuales**, gratuito y en web
  - **Concept Draw. Mind Map**, comercial
- Similitud de documentos
  - Detección de plagio y determinación de autoría
  - Poppins (Nazar, R.), clasificación temática de textos de economía, medicina e informática

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

❖ Bloque de explotación

- Alineadores
- Procesadores de corpus: etiquetadores morfosintácticos
- El etiquetado del corpus a este nivel suele comprender las fases de:
  - (1) segmentación (en el PLN este proceso recibe el nombre de 'tokenización', que consiste en la segmentación del texto en cadenas de caracteres y cifras que se encuentran entre espacios);
  - (2) lematización (especificación de la forma no marcada de cada palabra)
  - (3) análisis morfológico y categorial (asignación de posibles categorías gramaticales y morfológicas); y
  - (4) etiquetado morfosintáctico o desambiguación de las categorías gramaticales dudosas.

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Etiquetadores morfosintácticos

Diseño de corpus y software para terminología

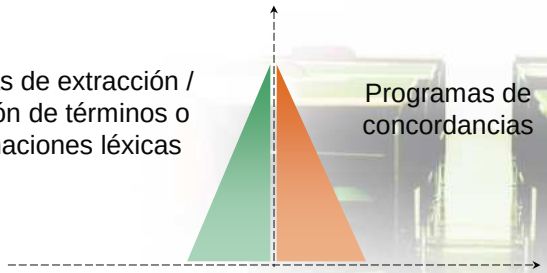
Contenido: Estación de trabajo terminológica y software

Nombre	Tipo	SO	Distribución	Lenguas
CLAWS (BNC)	Híbrido	Unix y Windows	Comercial, pero con versión online	Inglés
	http://ucrel.lancs.ac.uk/claws/			
Freeling (TALP, UPC)	Híbrido	Linux, Unix, Windows*	Gratuito, tb versión online	catalán, español, gallego, italiano, inglés
	http://www.lsi.upc.edu/~nlp/freeling/			
QTAG	Estadístico	Linux, Mac OSX y Windows	Gratuito	Independiente
	http://phrasys.net/uob/om/software			
TreeTagger	Estadístico	Linux, Mac OSX y Windows	Gratuito	Independiente
	http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html			

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Análisis del corpus



Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Análisis de corpus / sistemas de extracción/ detección de combinaciones (independientes)

1. ▪**Xtract (Smadja, 1993)**
2. ▪Termight (Dagan y Church, 1994)
3. ▪Champollion (Smadja et al., 1996)
4. ▪TRUCKS (Maynard, 1999)
5. YATE (Vivaldi, 2001)
6. Ngram Statistics Package (Pedersen & Banerjee, 2003)

Diseño de corpus y software para terminología

Análisis de corpus / sistemas de extracción/ detección de combinaciones (integrados o en web, con posibilidad de generar corpus ad hoc)

1. ▪Terminus, IULA
2. ▪Corpógrafo, Linguatec
3. ▪E-Termos
4. ▪Sketch Engine
5. WMatrix
6. …

Otra característica: Integran metodologías estándar de análisis de corpus: frecuencias, KWIC

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Programas de concordancias: Segmentan el corpus y ofrecen los datos en:

Diseño de corpus y software para terminología

Nombre	Aplicaciones	SO	Distribución
AntConc	concordancias, listados de palabras y palabras clave por frecuencia, listas de colocados y agrupaciones de una palabra base, gráfico de distribución	Linux y Windows	gratuita
	http://www.antlab.sci.waseda.ac.jp/antconc_index.html		
Conc	concordancias, índices y listados de palabras	Mac	gratuita
	http://www.sil.org/computing/conc/		
ConcApp	concordancias, listados de frecuencias	Windows	gratuita
	http://www.edict.com.hk/pub/concapp/		
Concordance	concordancias, listados de palabras y frecuencias, análisis de palabras clave	Windows	comercial
	http://www.concordancesoftware.co.uk/		
ConcGram	co-ocurrencias de una palabra (*congrams*), concordancias	Windows	comercial
	http://www.edict.com.hk/pub/concgram/		
KWIC Finder	concordancias de textos on line, agrupaciones léxicas con la aplicación *kfNgram*	Windows	gratuita
	http://www.kwicfinder.com/KWiCFinder.html		
Simple Concordance Program	Listados de palabras, concordancias, estadísticas	Windows y Mac	gratuita
	http://www.textworld.com/scp/		
TextSTAT	Listados de palabras y concordancias	Windows XP, Linux, Mac	gratuita
	http://neon.niederlandistik.fu-berlin.de/en/textstat/		
WordSmith Tools	estadísticas, listados mono y poliléxicos, concordancias, listado de palabras clave, gráfico de distribución y lista de colocados	Windows	comercial
	http://www.lexically.net/wordsmith/		

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

- **Las herramientas de extracción de términos pueden ser:**
  - **Monolingües**: analizan un texto o corpus para identificar candidatos a término.
  - **Bilingües**: analizan textos originales junto con sus traducciones a fin de identificar candidatos a término y sus equivalentes.

El vaciado terminológico es ASISTIDO y no COMPLETAMENTE AUTOMÁTICO

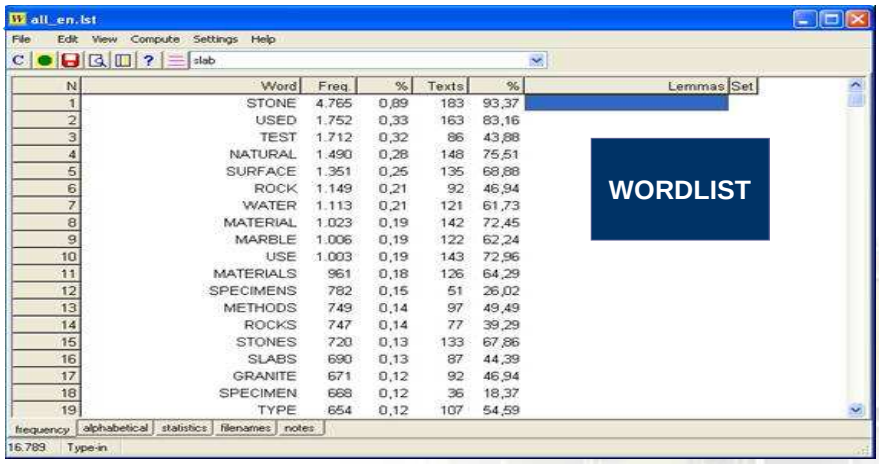
Contenido: Estación de trabajo terminológica y software

- **Hay tres métodos de extracción (Estopà, 1999):**
  - **Lingüístico**:
    - Intenta identificar combinaciones léxicos que coinciden con ciertos patrones morfosintácticos (e.g., ADJ+N or N+N).
    - Muy dependiente del idioma (los patrones de formación de términos difieren de una lengua a otra).
  - **Estadístico:**
    - Busca secuencias repetidas de unidades léxicas.
    - Independiente del idioma.
    - La cantidad de "ruido" es relativamente elevada.
  - **Híbrido:**
    - Ambos enfoques (lingüístico + estadístico).

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

Contenido: Estación de trabajo terminológica y software



N	Word	Freq	%	Texts	%	Lemmas	Set
1	STONE	4.765	0,89	183	93,37		
2	USED	1.752	0,33	163	83,16		
3	TEST	1.712	0,32	86	43,88		
4	NATURAL	1.490	0,28	148	75,51		
5	SURFACE	1.351	0,25	135	68,88		
6	ROCK	1.149	0,21	92	46,94		
7	WATER	1.113	0,21	121	61,73		
8	MATERIAL	1.023	0,19	142	72,45		
9	MARBLE	1.006	0,19	122	62,24		
10	USE	1.003	0,19	143	72,96		
11	MATERIALS	961	0,18	126	64,29		
12	SPECIMENS	782	0,15	51	26,02		
13	METHODS	749	0,14	97	49,49		
14	ROCKS	747	0,14	77	39,29		
15	STONES	720	0,13	133	67,86		
16	SLABS	690	0,13	87	44,39		
17	GRANITE	671	0,12	92	46,94		
18	SPECIMEN	668	0,12	36	18,37		
19	TYPE	654	0,12	107	54,59		

all_en_lematizada.lst

File Edit View Compute Settings Help

stone

N	Word	Freq.	%	Texts	%	Lemmas	Set
16	SLABS	690	0,13	87	44,39		
17	GRANITE	852	0,12	92	46,94	granite[671] granites[181]	
18	SPECIMEN	1.450	0,12	36	18,37	specimen[668] specimens[782]	
19	TYPE	1.021	0,12	107	54,59	type[654] types[367]	
20	DIFFERENT	643	0,12	131	66,84		
21	BUILDING	946	0,12	108	55,10	building[633] buildings[313]	
22	VARIETIES	610	0,11	47	23,98		
23	STRENGTH	614	0,11	84	42,86	strength[602] strengths[12]	
24	SIZE	770	0,11	119	60,71	size[600] sizes[170]	
25	STANDARD	739	0,11	90	45,92	standard[592] standards[147]	
26	RESISTANCE	567	0,11	90	45,92		
27	BLOCKS	561	0,10	79	40,31		
28	AESTHETIC	519	0,10	43	21,94	aesthetic[513] aesthetics[6]	
29	PRODUCTION	522	0,10	85	43,37	production[512] productions[10]	
30	COLOR	615	0,09	43	21,94	color[508] colors[107]	
31	PREN	505	0,09	28	14,29		
32	NUMBER	521	0,09	108	55,10	number[502] numbers[19]	
33	CUT	591	0,09	92	46,94	cut[501] cuts[90]	
34	MADE	492	0,09	130	66,33		

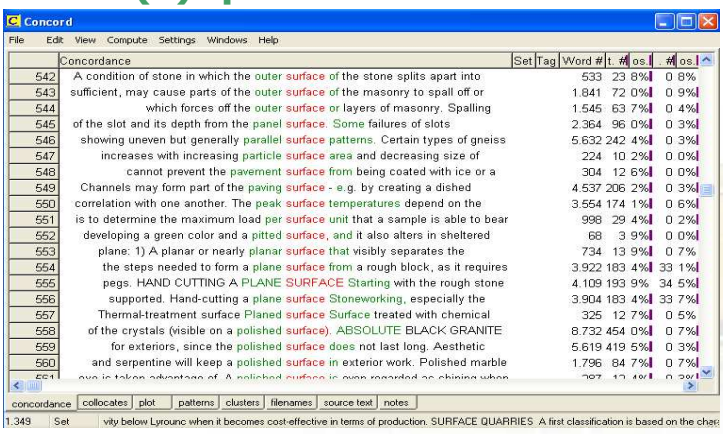
frequency alphabetical statistics filenames notes

16.789 Type-in SLABS

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

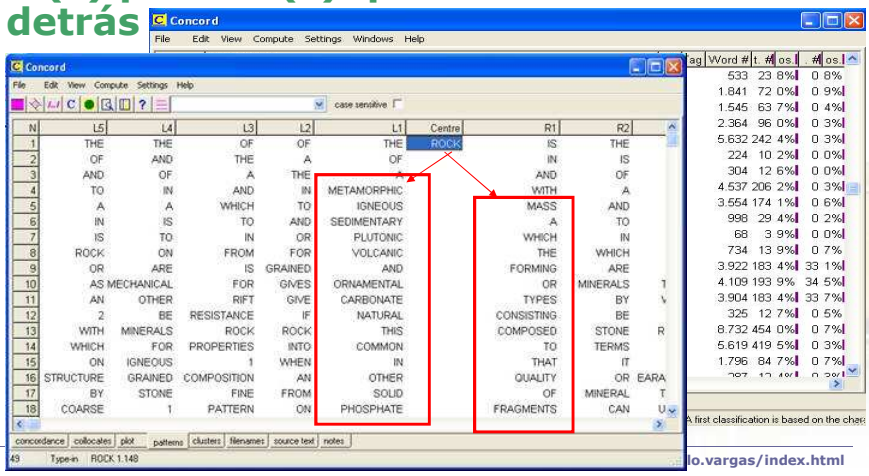
- **Pueden mostrar el término junto con la(s) palabra(s) que van delante o detrás**



Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

- **Pueden mostrar el término junto con la(s) palabra(s) que van delante o detrás**



Diseño de corpus y software para terminología

clasto

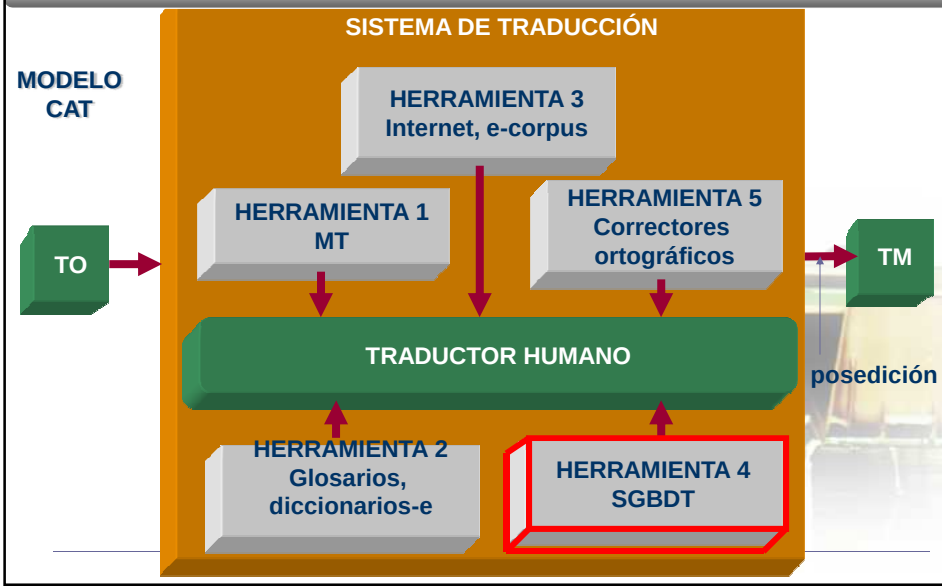
N	Concordance
1	o con matriz arcillosa en la que se ven algunos clastos. Accesorios: minerales de arcilla, opac
2	nas areniscas de grano fino-medio, con algunos clastos carbonatados y restos fósiles. Su colo
3	sicos. Es frecuente la presencia de zonas con clastos de dolomía que presentan una matriz m
4	e calcita. Mármol de calcita heterogranular con clastos redondeados de cuarzo. Minerales acc
5	arillo Espejón. Fig. 3.-Calizas lacustres con intraclastos y marmorización sinsedimentaria.
6	los clastos que presenten: ▪ Conglomerados, clastos > 2mm. ▪ Samitas, arenas consolida
7	o-medio que presenta una mayor proporción de bioclastos y granos de cuarzo. Son de color a
8	ma rojizo variable. Micrita con lamelibranquios e intraclastos de esparita. Accesorios: minerales
9	rta (Cehegín): roca carbonática con moluscos e intraclastos. Minerales accesorios: arcilla y cua
10	on las de grano grueso, siempre que no existan microclastos ni porfiroblastos, es decir, siempr
11	tas) cementados por calcita espática. Tanto los clastos como el cemento contienen los óxido
12	de cuarzo y feldespatos del 15 %, estando los bioclastos constituidos por foraminíferos, globi
13	cas carbonatadas ignorando el tamaño de los clastos, de los materiales amorfos o de los org
14	samitas se clasifican por la composición de los clastos de mayor tamaño que componen el esq
15	e conocidas como areniscas, si predominan los clastos entre 2 mm y 0,06 mm. ▪ Lutitas, si l
16	tipo caliche que puede acabar englobando los clastos. II.- Rocas metamórficas Las rocas m
17	or tamaño que componen el esqueleto, y de los clastos menores que forman la matriz, así com
18	ificaciones, descritas como cataclásticas si los clastos no están alargados y como milonítica
19	calcita microesparítica de grano muy fino; los bioclastos, que suponen entre el 50 y 60% de l
20	sificar en primera instancia por el tamaño de los clastos que presenten: ▪ Conglomerados, cla
21	ente, los procesos diagenéticos cementarán los clastos resultantes de la erosión y el transporte
22	astos entre 2 mm y 0,06 mm. ▪ Lutitas, si los clastos son menores de 0,06 mm; si los tamañ
23	míticas de Linares-huero". Está constituida por bioclastos (equínidos, briozoos, foraminíferos,
24	bioesparita, constituida en más de un 50% por bioclastos y con proporción Esparita/Micrita d
25	) placa, 137 del Museo de Ciencias presentan intraclastos de gran tamaño (1 a 6 cm) y en alg
26	calclititas, ya que en casi todos los casos son exoclastos). Conglomerados (5,4%): marmoriza
27	erados (fig.2), placa 131 del M. de Ciencias son clasto- soportados y con frecuentes contactos
28	acterizando principalmente la morfología de sus clastos y la composición mineralógica de los s
29	nitas, tobas pozolánicas, coladas basálticas y piroclastos soldados constituyen los principales

Chelo Vargas - http://www.ua.es/personal/chelo.vargas/index.html

	Word 1	Freq	Word 2	Freq.	Text	Gap	Joint	MI	Z	MI3	Log L.
5935	CALIZA	249	ORNAMENTAL	137	4	1	7	6,93	8,49	-9,17	32,97
5939	CALIZA	249	GRIS	179	4	1	7	6,55	7,22	-9,17	11,5
5945	CALIZA	249	AZUL	44	1	1	6	8,35	13,55	-9,84	158,3
5947	CALIZA	249	COMPACTA	50	5	1	7	8,39	14,84	-9,17	144,53
5951	CALIZA	249	PACKSTONE	5	1	1	4	10,9	27,5	-11,6	302,94
5952	CALIZA	249	CRISTALINA	50	3	1	3	7,16	6,1	-12,84	144,53
5954	CALIZA	249	BLANCA	34	3	1	7	8,94	18,17	-9,17	184,48
5956	CALIZA	249	BLANDA	22	3	1	3	8,35	9,58	-12,84	223,03
5962	CALIZA	249	FOSILÍFERA	8	2	1	4	10,22	21,67	-11,6	285,02
5963	CALIZA	249	DOLOMITIZADA	5	3	1	5	11,22	34,41	-10,63	302,94
5967	CALIZA	249	DOLOMÍTICA	9	1	1	4	10,05	20,41	-11,6	279,58
5968	CALIZA	249	BRECHOIDE	15	1	1	4	9,32	15,71	-11,6	250,81
5971	CALIZA	249	ARRECIFAL	5	1	1	5	11,22	34,41	-10,63	302,94
5972	CALIZA	249	MARMÓREA	11	7	1	8	10,76	37,06	-8,6	269,33
5973	CALIZA	249	PORTLAND	21	1	1	4	8,83	13,19	-11,6	226,71
5975	CALIZA	249	FRANCESA	10	2	1	5	10,22	24,23	-10,63	274,36
5976	CALIZA	249	IRLANDESA	6	1	1	3	10,22	18,77	-12,84	296,65
5977	CALIZA	249	CRETÁCICA	9	1	1	5	10,37	25,56	-10,63	279,58
5984	CALIZAS	314	ORNAMENTALES	253	11	1	38	8,15	31,72	-1,85	6,58
6000	CALIZAS	314	GRIS	179	2	1	3	4,99	2,11	-12,84	37,44
6007	CALIZAS	314	COMPACTAS	31	5	1	7	8,74	16,9	-9,17	269,75
6012	CALIZAS	314	DOLOMÍTICAS	13	3	1	3	8,77	11,19	-12,84	343,99
6013	CALIZAS	314	LACUSTRES	10	1	1	4	9,57	17,18	-11,6	359,91
6014	CALIZAS	314	BIOCLÁSTICAS	5	3	1	3	10,15	18,3	-12,84	390,75
6015	CALIZAS	314	MARMÓREAS	20	8	1	19	10,81	58,14	-4,85	311,63
6016	CALIZAS	314	RECRISTALIZADAS	8	2	1	5	10,21	24,12	-10,63	371,47
6021	CALIZAS	314	ARENOSAS	9	1	1	6	10,3	27,3	-9,84	365,58
6022	CALIZAS	314	FOSILÍFERAS	8	3	1	7	10,69	33,85	-9,17	371,47
6026	CALIZAS	314	OOLÍTICAS	5	3	1	5	10,89	30,61	-10,63	390,75
6027	CALIZAS	314	CRETÁCICAS	10	4	1	9	10,74	38,93	-8,09	359,91
6028	CALIZAS	314	JURÁSICAS	8	2	1	6	10,47	28,98	-9,84	371,47
6029	CALIZAS	314	CONTINENTALES	7	2	1	4	10,08	20,61	-11,6	377,6
6030	CALIZAS	314	TERCIARIAS	12	2	1	8	10,3	31,53	-8,6	349,13
6031	CALIZAS	314	MICRÍTICAS	8	2	1	4	9,89	19,25	-11,6	371,47
6032	CALIZAS	314	MARMORIZADAS	9	1	1	7	10,52	31,89	-9,17	365,58

Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software



Diseño de corpus y software para terminología

Contenido: Estación de trabajo terminológica y software

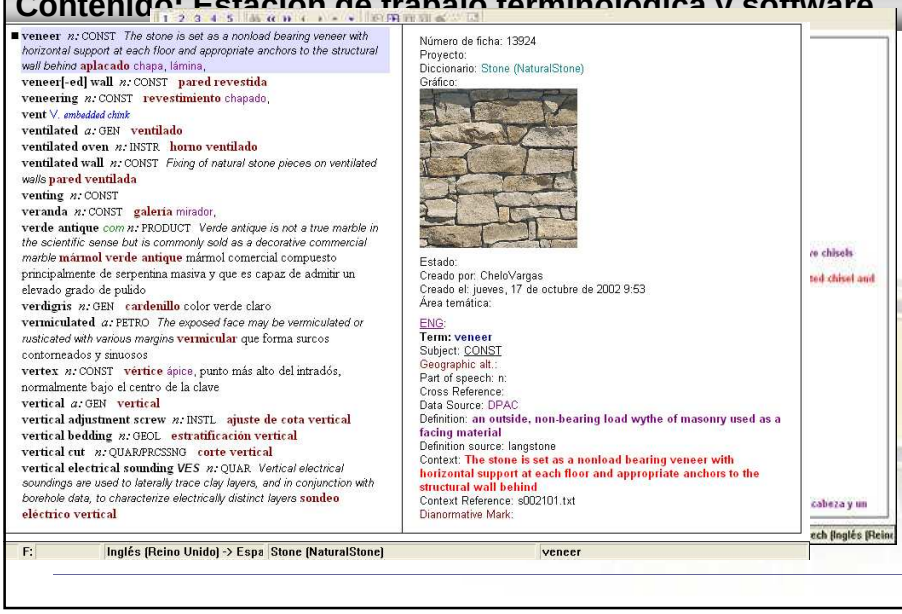
- **Un SMT puede ayudar en algunas tareas terminológicas del traductor, como:**
  - almacenamiento
  - recuperación
  - actualización de las entradas terminológicas.
- **En la actualida, todos los SMT incluyen SGBDT**

Diseño de corpus y software para terminología

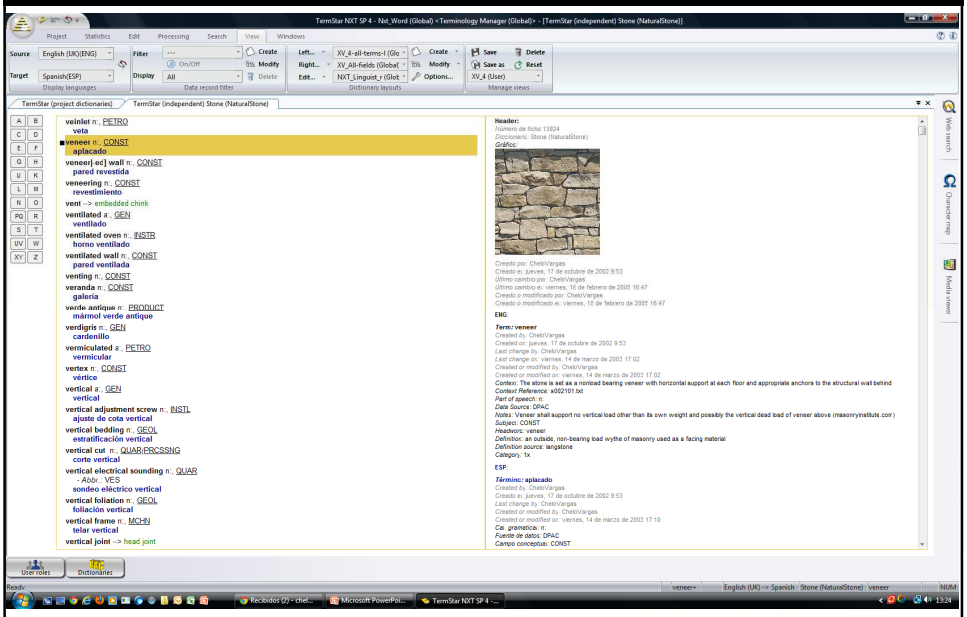
Contenido: Estación de trabajo terminológica y software

Diseño de corpus y software para terminología

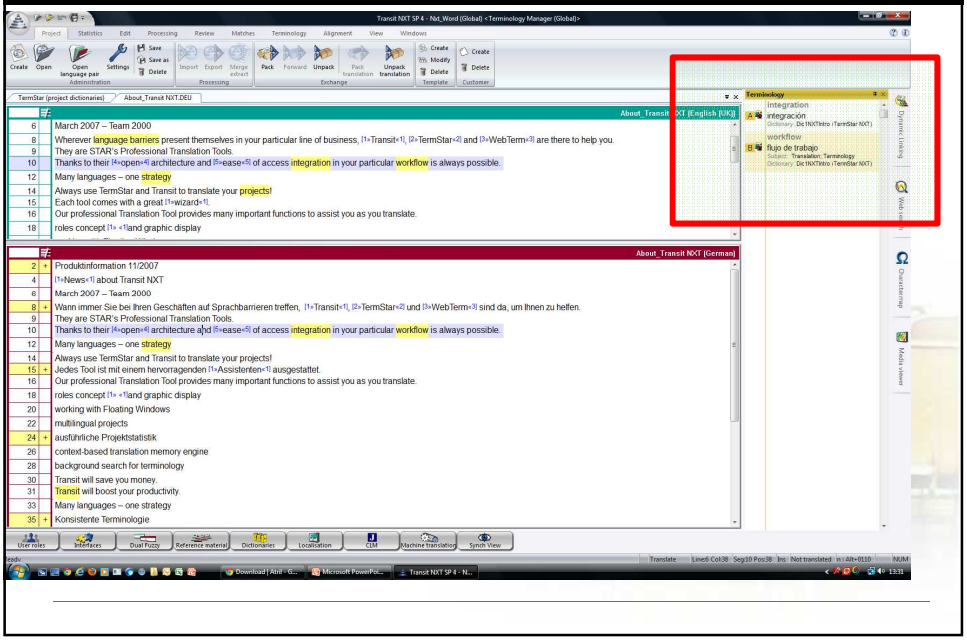
Contenido: Estación de trabajo terminológica y software



Diseño de corpus y software para terminología



Diseño de corpus y software para terminología



Diseño de corpus y software para terminología

Contenid: PUNTOS CLAVE

- Las contribuciones de la informática han supuesto cambios en la metodología terminológica y en el procesamiento de los términos.
- Hay 5 etapas básicas en terminografía en los que la informática representa un papel primordial.
- Cuando se diseña un corpus, el usuario debe determinar criterios precisos analizando los fines de su proyecto.
- Hay 4 grandes bloques de trabajo en una ETT: investigación, desarrollo, explotación y gestión
- En cada uno de estos bloques hay herramientas informáticas que nos ayudan a hacer tareas específicas

Diseño de corpus y software para terminología

Contenid: PUNTOS CLAVE

- El vaciado terminológico es semiautomático y no completamente automático
- Hay 3 tipos de extractores terminológicos: lingüísticos, estadísticos e híbridos.
- SMT pueden ayudar al traductor en tareas terminológicas.
- SGBDT se incluyen en todos los SMT