



RECIDA

Red de Centros de Información y Documentación Ambiental

Taller práctico: Crear y visualizar nuevo conocimiento extraído de un corpus documental mediante Text Mining

Tag cloud del perfil investigador del
Dr. Jesús Tramullas

RECIDA 2013

Dr. José Pino Díaz

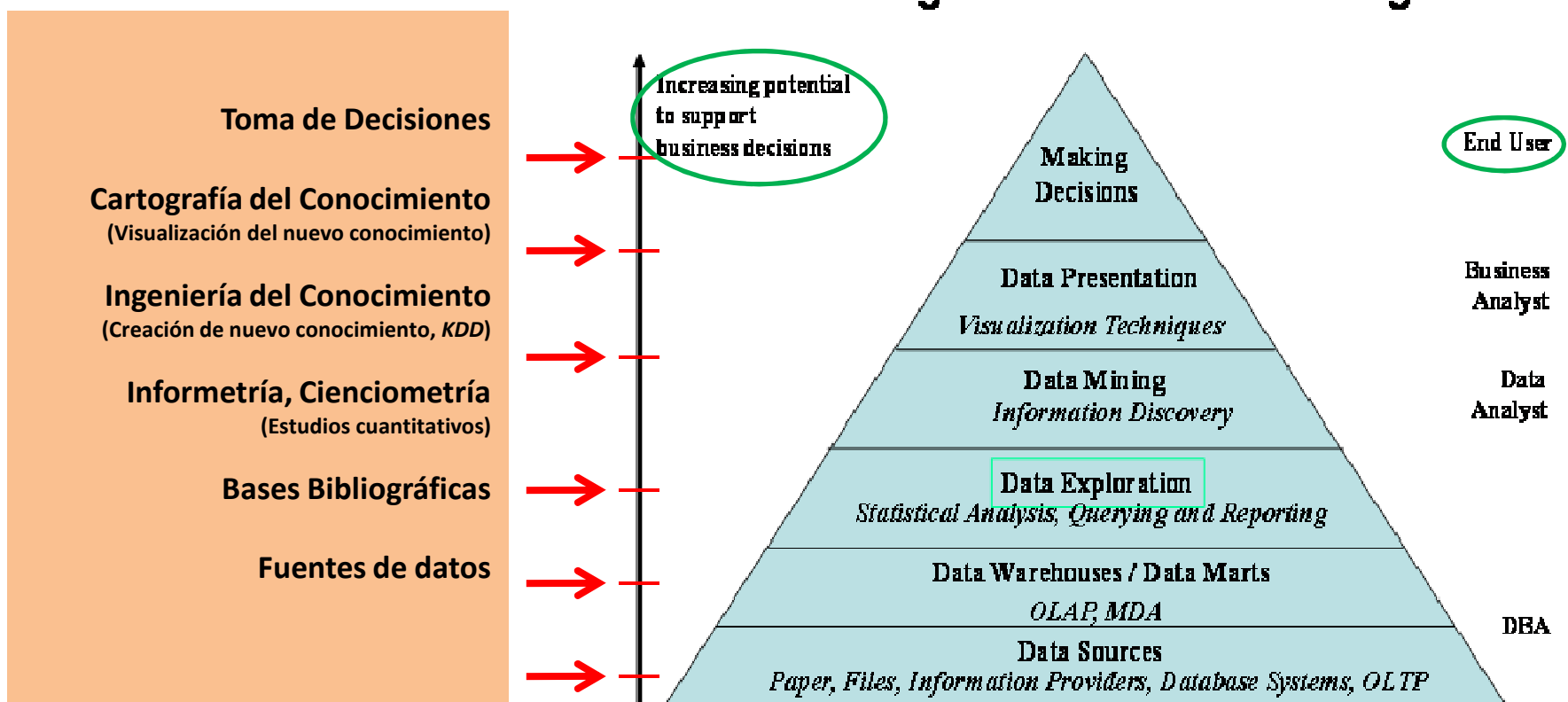
Valsaín (Segovia), 17-19 junio.



RECIDA

Red de Centros de Información y Documentación Ambiental

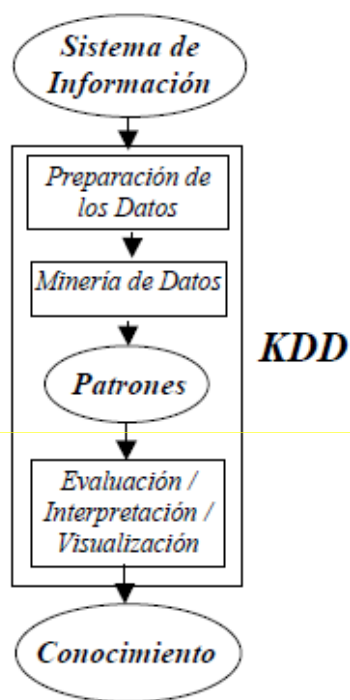
Data Mining and Business Intelligence



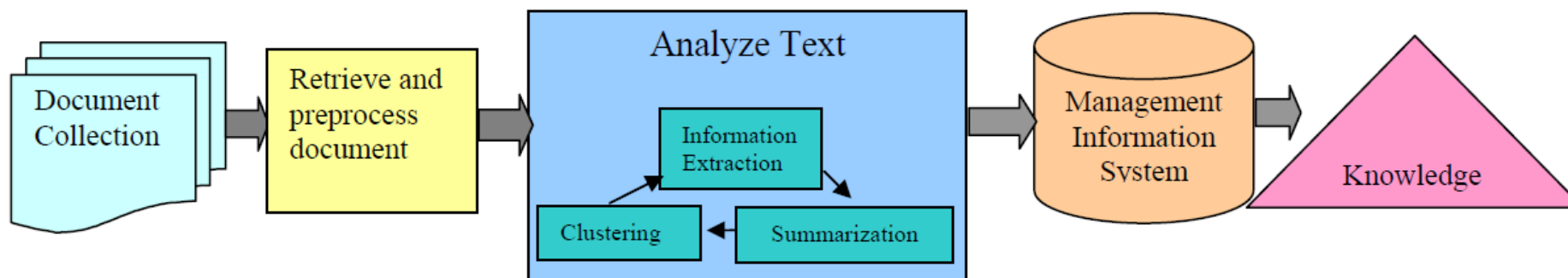


RECIDA

Red de Centros de Información y Documentación Ambiental



1. Determinar las fuentes de información que pueden ser útiles y dónde conseguirlas.
2. Diseñar el esquema de un almacén de datos (Data Warehouse) que consiga unificar de manera operativa toda la información recogida.
3. Implantación del almacén de datos que permita la “navegación” y visualización previa de sus datos, para discernir qué aspectos puede interesar que sean estudiados.
4. Selección, limpieza y transformación de los datos que se van a analizar. La selección incluye tanto una criba o fusión horizontal (filas) como vertical (atributos).
5. Seleccionar y aplicar el método de minería de datos apropiado.
6. Interpretación, transformación y representación de los patrones extraídos.
7. Difusión y uso del nuevo conocimiento.





RECIDA

Red de Centros de Información y Documentación Ambiental

Text mining applied

	information extraction	topic tracking	summarization	categorization	clustering	concept linkage	information visualization	question answering
Medical:								
FAQ's	x			x		x		x
Drug design	x				x	x		
New treatment		x				x		
Business:								
Competitive Analysis		x	x					
Media impact / analysis		x						
Current Awareness		x						
Intellectual property infringement	x	x			x			
Customer support for FAQ's	x			x	x			x
Social network detection							x	
Content personalization		x			x			
Government:								
Homeland security: detecting terrorist networks	x	x			x	x	x	
Law enforcement: crime detection / prevention	x	x			x	x	x	
Education:								
Research on a topic		x	x	x				
Citation analysis	x				x		x	
FAQ's	x			x	x			x

Table 3. Some examples of where text mining tools can be applied to the fields of medicine, business, government, and education.



RECIDA

Red de Centros de Información y Documentación Ambiental

CCHS ICYT

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACION

CSIC

CENTRO DE CIENCIAS HUMANAS Y SOCIALES

ISOC en cifras Vocabulario ISOC Documentos FAQs Contacto Sugerencias Información para editores

La Base de Datos ISOC recopila y difunde desde 1976 la producción científica del área de las Ciencias Humanas y Sociales publicada en revistas españolas. Esta base de datos multidisciplinar cubre las siguientes áreas temáticas: Antropología, Arqueología y Prehistoria, Arte, Biblioteconomía y Documentación, Ciencias Políticas, Derecho, Economía, Educación, Filosofía, Geografía, Historia, Lingüística y Literatura, Psicología, Sociología, Urbanismo y Estudios sobre América Latina.

Idiomas

-  English
-  Español

Enlaces

- Base de Datos ISOC por suscripción
- Sumarios ISOC
- ISOC - Biblioteconomía y Documentación
- Directorio 'Revistas de Ciencias Sociales y Humanidades'
- Tesoros ISOC
- Vocabulario ISOC

ISOC en cifras (8 de enero de 2013)

- **678.363 referencias bibliográficas** (el 95 % corresponden a artículos publicados en revistas)
- **351.856 registros con afiliación institucional de los autores**
- **245.375 registros con resumen de autor**
- **165.536 enlaces al texto completo de artículos disponibles en acceso abierto**
- **Crecimiento anual de 30.000 referencias (en los últimos 4 años de producción)**
- **953.456 accesos por suscripción (año 2012)**
- **2.764 revistas en total. La selección actual es de 1.357 títulos, de los cuales 943 tienen versión electrónica.**

El Grupo de Investigación de Evaluación de las Publicaciones Científicas del IEDCYT (CSIC) realiza una valoración de las revistas que determina su aceptación o no en ISOC. Para conocer el sistema de selección consulte el apartado [Información para editores](#).

ISOC es la principal base de datos de la producción científica española en Ciencias Humanas y Sociales.

A 8 de enero de 2013 tiene cerca de 680.000 referencias bibliográficas





RECIDA

Red de Centros de Información y Documentación Ambiental

informe_docu_tramullas.txt - Bloc de notas

Archivo Edición Formato Ver Ayuda

Salida de documentos de ICYT - Ciencia y Tecnología:
Parámetros de búsqueda: Autores= "tramullas"

Documento nº 1
Autores.....: Tramullas, J.
Titulo Revista.....: Novática
Descriptores.....: Sistemas de información automatizados; Documentación
CopyRight.....: © CSIC. Base de Datos ICYT. Todos los derechos reservados.

Salida de documentos de ISOC - Ciencias Sociales y Humanidades:
Parámetros de búsqueda: Autores = "tramullas jesús"

Documento nº 1
Autores.....: Tramullas, Jesús
Titulo Revista.....: Anuario ThinkEPI
CopyRight.....: © CSIC. Base de Datos ISOC. Todos los derechos reservados.

Documento nº 2
Autores.....: Tramullas, Jesús
Titulo Revista.....: Anuario ThinkEPI
Descriptores.....: Internet; World Wide Web
CopyRight.....: © CSIC. Base de Datos ISOC. Todos los derechos reservados.

Documento nº 3
Autores.....: Tramullas Saz, Jesús; Garrido Picazo, Piedad
Titulo Revista.....: Boletín de la ANABAD
Descriptores.....: Bibliotecas; Bibliotecas digitales
CopyRight.....: © CSIC. Base de Datos ISOC. Todos los derechos reservados.

Documento nº 4
Autores.....: Tramullas, Jesús
Titulo Revista.....: Anuario ThinkEPI
Descriptores.....: Correo electrónico
CopyRight.....: © CSIC. Base de Datos ISOC. Todos los derechos reservados.



RECIDA

Red de Centros de Información y Documentación Ambiental

informe_docu_mayúsculas.txt - Bloc de notas
Archivo Edición Formato Ver Ayuda Maximiz

NOVATICA,SISTEMAS-DE-INFORMACION-
AUTOMATIZADOS,DOCUMENTACION,
ANUARIO-THINKEPI,
ANUARIO-THINKEPI,INTERNET,WORLD-WIDE-WEB,
GARRIDO-PICAZO-PIEDAD,BOLETIN-DE-LA-
ANABAD,BIBLIOTECAS,BIBLIOTECAS-DIGITALES,
ANUARIO-THINKEPI,CORREO-ELECTRONICO,
EL-PROFESIONAL-DE-LA-INFORMACION,WORLD-WIDE-
WEB,ACCESIBILIDAD,ESTRUCTURA-DE-LA-INFORMACION,
ANUARIO-THINKEPI,INTERNET,PROCESAMIENTO-DE-LA-
INFORMACION,NUEVAS-TECNOLOGIAS,DIGITALIZACION,BUSCADORES,
SANCHEZ-CASABON-ANA-I.,GARRIDO-PICAZO-PIEDAD,EL-PROFESIONAL-
DE-LA-INFORMACION,
BIBLIOTECAS,BIBLIOTECAS-DIGITALES,SERVICIOS-DE-ORIENTACION-
BIBLIOGRAFICA,CONSULTAS-BIBLIOGRAFICAS,
ANUARIO-THINKEPI,WORLD-WIDE-WEB,GESTION-DE-LA-INFORMACION,
ANUARIO-THINKEPI,WORLD-WIDE-WEB,BUSQUEDA-DE-
INFORMACION,RECUPERACION-DE-INFORMACION,
GARRIDO-PIEDAD,MARTINEZ-FRANCISCO,FUERTES-GABRIEL,RIED-
REVISTA-IBEROAMERICANA-DE-EDUCACION-A-DISTANCIA,
GARRIDO-PICAZO-PIEDAD,EL-PROFESIONAL-DE-LA-
INFORMACION,WORLD-WIDE-WEB,ARCHIVOS-ABIERTOS,EVALUACION-DE-
-COLECCIONES,SOFTWARE
GARRIDO-PICAZO-PIEDAD,
GARRIDO-PICAZO-PIEDAD,
HIPERTEXT.NET,GESTION-DE-LA-INFORMACION,SOFTWARE,PROGRAMAS-
INFORMATICOS,
NAVARRO-BONILLA-DIEGO,REVISTA-ESPAÑOLA-DE-DOCUMENTACION-
CIENTIFICA,DIRECTORIOS-DE-RECURSOS,INTERNET,CLASIFICACION,
EL-PROFESIONAL-DE-LA-INFORMACION,WORLD-WIDE-WEB,ESTRUCTURA
-DE-LA-INFORMACION,SITIOS-WEB,
GARRIDO-PICAZO-PIEDAD,EL-PROFESIONAL-DE-LA-
INFORMACION,BIBLIOTECAS-DIGITALES,INTERFACES,DESARROLLO-DE-
COLECCIONES,SOFTWARE,
FORINF@-REVISTA-IBEROAMERICANA-DE-LOS-USUARIOS-DE-LA-
INFORMACION,
BIBLIOTECAS,BIBLIOTECAS-UNIVERSITARIAS,BIBLIOTECAS-DIGITALES,
RECURSOS-ELECTRONICOS,RECURSOS-DE-INFORMACION,PORTALES,
EL-PROFESIONAL-DE-LA-INFORMACION,GESTION-DEL-CONOCIMIENTO,
DISEÑO-GRAFICO,INFORMACION,DOCUMENTOS-ELECTRONICOS,



RECIDA

Red de Centros de Información y Documentación Ambiental

www.wordle.net

Wordle™

[Home](#)

[Create](#)

[Gallery](#)

[Credits](#)

[News](#)

[Forum](#)

[FAQ](#)

[Advanced](#)

Wordle is a toy for generating “word clouds” from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text. You can tweak your clouds with different fonts, layouts, and color schemes. The images you create with Wordle are yours to use however you like. You can print them out, or save them to the Wordle gallery to share with your friends.

[Create your own.](#)

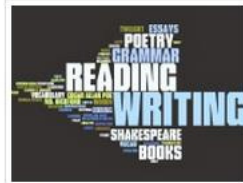
View some examples created by others...



[English notebook cover](#)

by Ace Acedemic!

3 years, 10 months ago



[Period G](#)

by Meredith

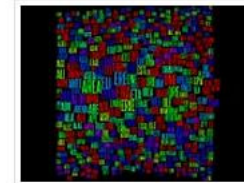
3 years, 9 months ago



[US Constitution](#)

by Jonathan

3 years, 6 months ago



[Most Common Crossword](#)

[Answers](#)

by Jonathan

4 years, 6 months ago

[More...](#)

© 2013 [Jonathan Feinberg](#)

[Terms of Use](#)

build #1375



RECIDA

Red de Centros de Información y Documentación Ambiental

Wordle™

[Casa](#)

[Crear](#)

[Galería](#)

[Créditos](#)

[Noticias](#)

[Foro](#)

[Preguntas
más
frecuentes](#)

Pegar en un montón de texto:

O

Introduce la URL de cualquier blog, blog feed o cualquier otra página web que tiene un feed Atom o RSS.



RECIDA

Red de Centros de Información y Documentación Ambiental

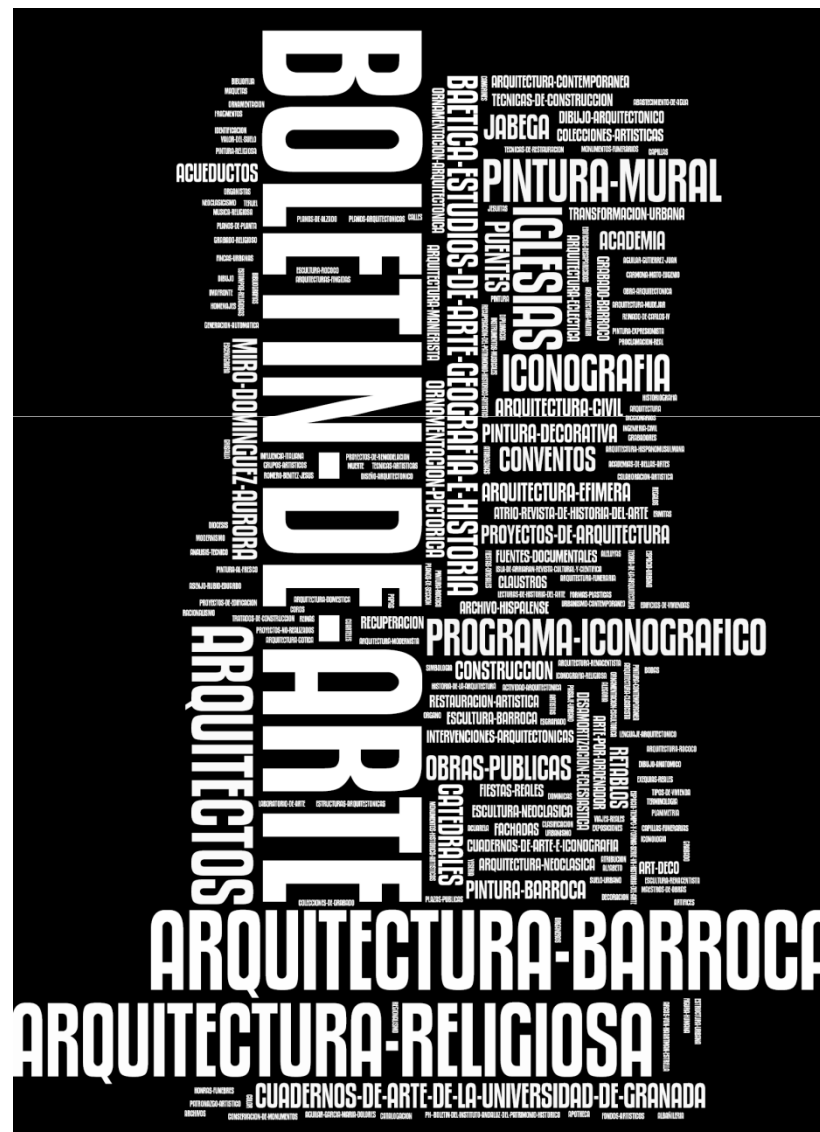
Ejemplos



RECIDA

Red de Centros de Información y Documentación Ambiental

Perfil investigador de la Dra. Rosario Camacho, según ISOC





RECIDA

Red de Centros de Información y Documentación Ambiental

Muchas Gracias

Prof. Dr. José Pino Díaz



DEPARTAMENTO DE HISTORIA DEL ARTE
Universidad de Málaga

Páginas personales: uma.academia.edu/JosePinoDiaz

researchgate.net/profile/Jose_Pino-Diaz/

Grupo Investigación Techné: techne-research-group.com/

Blog Áreadoc: areadoc.blogspot.com

Facebook: facebook.com/jose.pino.diaz

Twitter: [@jpinod02](https://twitter.com/jpinod02)

Linkedin: es.linkedin.com/in/josepinodiaz/

Competitive intelligence: ci2020.com/profile/JosePinoDiaz

Slides: slideshare.net/Jose_Pino/slideshows

Scribd: scribd.com/Jose_Pino

Screencast: screencast.com/users/Areadoc/

Youtube: www.youtube.com/user/JosePinoDiaz