

Text Mining

para la

toma de decisiones

Curso "Experto Universitario en Gestión de Recursos Humanos en Entornos Digitales"

<http://www.recursoshumanosonline.uma.es/>

Prof. Dr. José Pino-Díaz

Departamento de Economía y Administración de Empresas

Universidad de Málaga

Málaga, 8 de febrero de 2019

Text Mining para la toma de decisiones

1. ¿Qué es *Text Mining*?
2. Proceso de creación de nuevo conocimiento (KDD)
3. Herramientas de *Text Mining* y campos disciplinares de aplicación
4. Aplicaciones de *Text Mining*
5. Técnicas de *HR Analytics* en Recursos Humanos
6. Ejemplo 1: Aplicación de minería de textos en bases de datos bibliográficas (empleo de Rstudio, Bibliometrix y Biblioshyni).
7. Ejemplo 2: Aplicación de minería de textos en análisis de respuestas en encuestas abiertas (empleo de R.TeMiS).

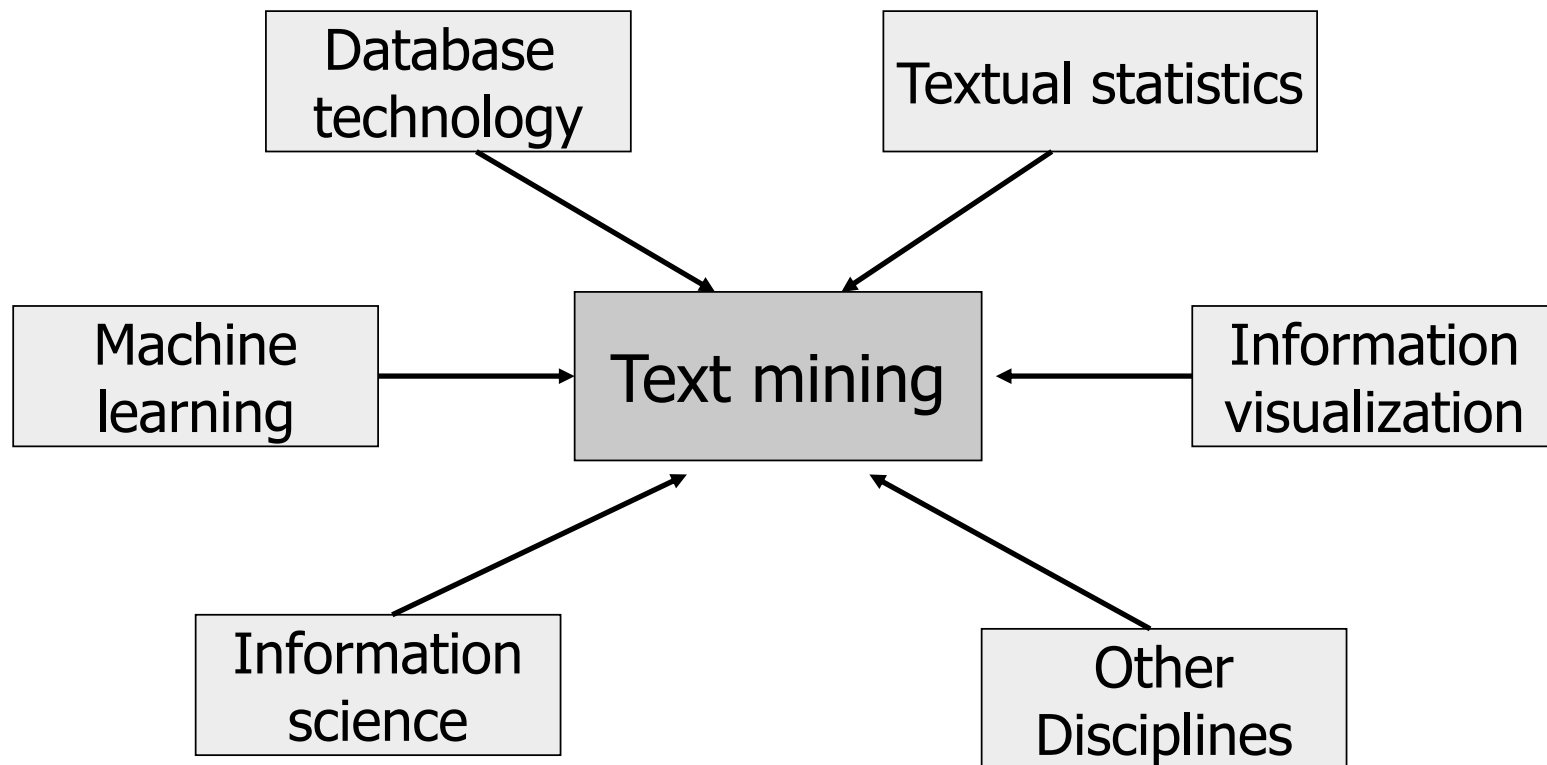
¿Qué es *Text Mining*?

La minería de textos es un área de la informática relacionada con el procesamiento del lenguaje natural, la minería de datos, el aprendizaje automático, la recuperación de información y la gestión del conocimiento.

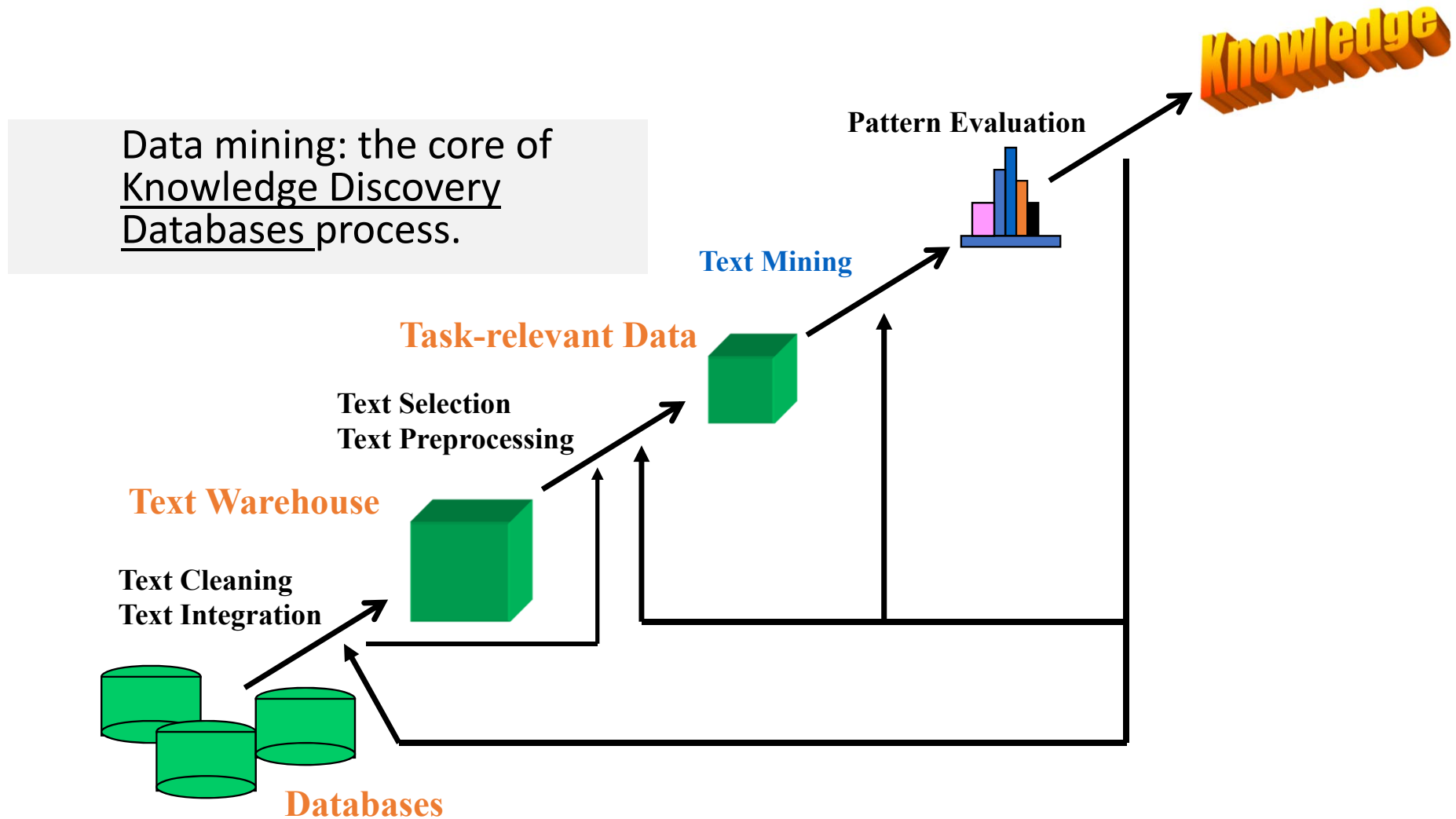
La minería de textos se emplea para extraer información útil de datos textuales no estructurados a través de la identificación y exploración de patrones interesantes¹.

1 Bharati N Kharada en "Text Mining and its applications, Academia.edu; recuperado el 1 de febrero de 2019, disponible en http://www.academia.edu/6888464/Text_Mining_and_Its_Applications, citando a Hotho A, Nürnberger A, Paaß G & AiS F (2005). A Brief Survey of Text Mining, recuperado el 1 de febrero de 2019, disponible en <https://pdfs.semanticscholar.org/9643/0cc91ed91fd2d4042fa6fcb7ecf4005d77a7.pdf>

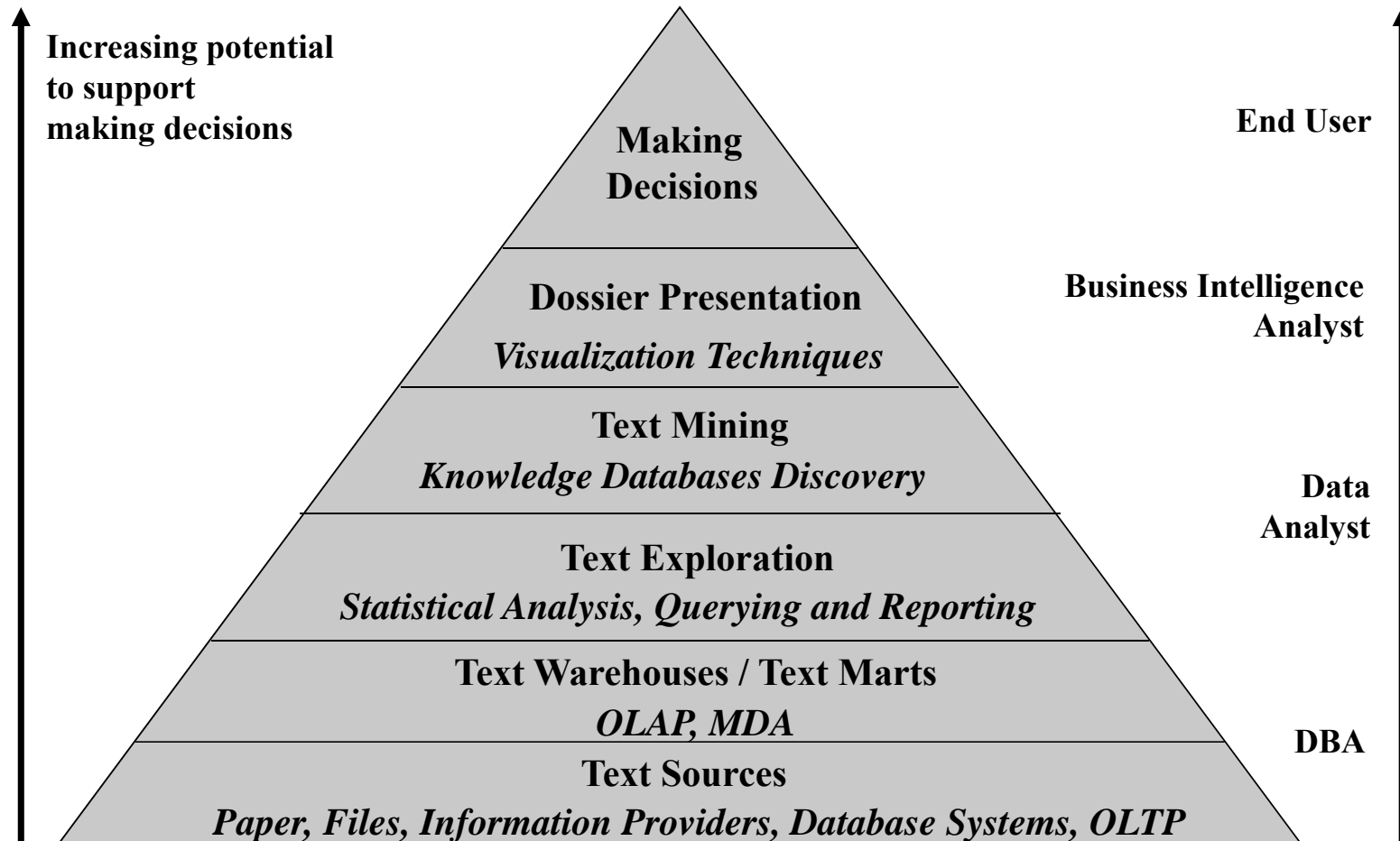
Text Mining, confluencia de múltiples disciplinas



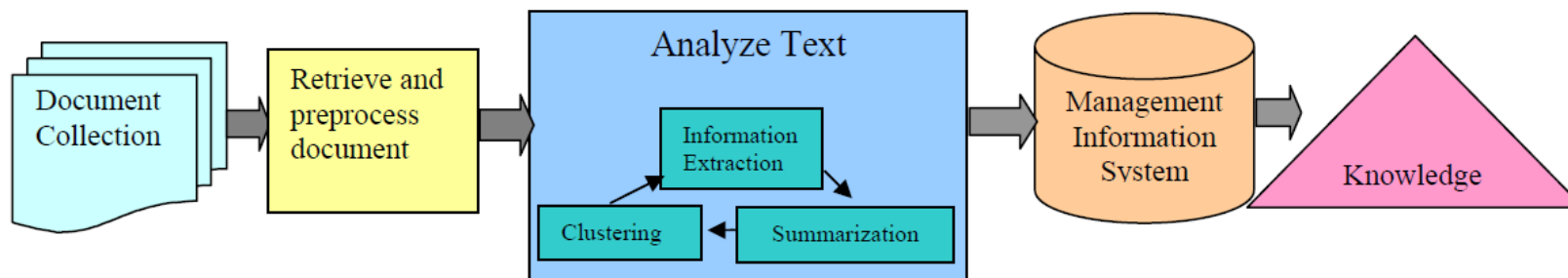
Text Mining, un proceso de creación de nuevo conocimiento (KDD)



Text Mining y Toma de decisiones



Text Mining, un proceso que da valor añadido



En 2001, Dow Chemicals se unió a Union Carbide Corporation (UCC). Esta unión requirió la integración de 35.000 informes de UCC en el sistema de gestión del documentos de Dow. Dow Chemicals eligió a la empresa ClearForest, líder en *text-driven business solutions*, para realizar la integración de la colección de documentos. Usando tecnología ad-hoc se identificaron sustancias químicas, productos, empresas, centros y personas. Esto permitió que Dow Chemicals agregara más de 80 años de investigación de UCC a su sistema de gestión de información y aproximadamente 100.000 nuevas sustancias químicas a su registro. Cuando el proyecto se completó, se estimó que Dow Chemicals ahorró casi 3 millones de dólares y que el tiempo empleado en clasificar los documentos se redujo un 50% y los errores de datos entre un 10-15%¹.

1 Fan W, Wallace L, Rich S, Zhang Z (2005). "Tapping Into the Power of Text Mining"; recuperado el 1 de febrero de 2019, disponible en <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=64E863EEAA4C8B00E627E98E4B74707D?doi=10.1.1.81.1579&rep=rep1&type=pdf>

Herramientas de *Text Mining* y campos de aplicación

	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.

1 Fan W, Wallace L, Rich S, Zhang Z (2005). "Tapping Into the Power of Text Mining"; recuperado el 1 de febrero de 2019, disponible en <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=64E863EEAA4C8B00E627E98E4B74707D?doi=10.1.1.81.1579&rep=rep1&type=pdf>

Text Mining, ejemplo de uso de la herramienta *Concept Linkage*

Una aplicación muy popular del *Text Mining* es relatada en *Hearst (Untangling Text Data Mining, 1999)*, Don Swanson intenta extraer información derivada de colecciones de texto. Teniendo en cuenta que los expertos sólo pueden leer una pequeña parte de lo que se publica en su campo, por lo general no se dan cuenta de los nuevos desarrollos que se suceden en otros campos¹.

Así, Swanson demostró cómo cadenas de implicaciones causales dentro de la literatura médica pueden conducir a hipótesis para enfermedades poco frecuentes, algunas de las cuales han recibido pruebas de soporte experimental. Investigando las causas de la migraña, dicho investigador extrajo evidencias a partir de títulos de artículos presentes en la literatura biomédica. Estas evidencias le sugirieron que la deficiencia de magnesio podría representar un papel en algunos tipos de migraña, una hipótesis que no existía en la literatura y que Swanson encontró mediante esas relaciones. De acuerdo con Swanson (Swanson y otros, 1994), estudios posteriores han probado experimentalmente esta hipótesis obtenida por *Text Mining* con buenos resultados.



1 “Data mining: torturando a los datos hasta que confiesen” en UOC; recuperado el 1 de febrero de 2019, disponible en <https://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>

Text Mining y sus aplicaciones¹

1. Análisis de respuestas de encuestas abiertas.
2. Procesamiento automático de mensajes y correos electrónicos.
3. Análisis de reclamaciones de garantía o seguros, entrevistas de diagnóstico, etc.
4. Investigar competidores rastreando sus sitios web.
5. Mejora de la búsqueda web.
6. Análisis de patentes.
7. Clasificación de textos para agencias de noticias.
8. Bioinformática.
9. Filtrado Anti-Spam de Correos Electrónicos.
10. Minería de textos en bases de datos bibliográficas.

¹ Bharati N Kharada en "Text Mining and its applications, Academia.edu; recuperado el 1 de febrero de 2019, disponible en http://www.academia.edu/6888464/Text_Mining_and_Its_Applications

Técnicas de *HR Analytics* en Recursos Humanos

Talent Analytics

People Analytics

HR Analytics

Diferentes conceptos para definir el conjunto de acciones que consisten en “*utilizar los datos ... para mejorar la toma de decisiones*”¹ (Beatriz Lucía, Directora de *Talent Analytics* en el Instituto de Ingeniería del Conocimiento) .

1 “De la evaluación psicológica al Talent Analytics” en Blog del Instituto de Ingeniería del Conocimiento (iic); recuperado el 1 de febrero de 2019, disponible en <http://www.iic.uam.es/rr-hh/evaluacion-psicologica-talent-analytics/>

Text Mining y Toma de decisiones en Recursos Humanos

Ejemplo 1. Minería de textos en SCOPUS sobre Text Mining y HR Analytics.



Ejemplo 2. Análisis de respuestas de encuestas abiertas en un MOOC para empleados públicos.



Text Mining con R

Módulo 1. Introducción: R software con licencia libre.



R es el software estadístico con licencia libre más extendido a nivel mundial tanto para docencia como para investigación.

Actualmente (08/02/2019) se encuentran disponibles más de **13.691 paquetes** desarrollados en R, que cubren multitud de campos científicos y tecnológicos.

Toda la información sobre R está disponible en la web <https://www.r-project.org/>

Existen interfaces gráficas que facilitan el uso de R: **RStudio**, **RKward** y **R Commander** se encuentran entre ellas.

Recursos docentes del módulo



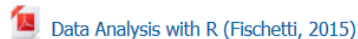
The R Project for Statistical Computing

Sitio web del Proyecto R para Estadística Computacional (**The R Project for Statistical Computing**).



Videotutoriales de R y R Commander

Recursos complementarios



Data Analysis with R (Fischetti, 2015)

Data Analysis with R (Fischetti, 2015). Reciente monografía en inglés de Tony Fischetti sobre análisis de datos con R, editado por Packt Publishing (primera edición 22 de diciembre de 2015).



R for Data Science (Dan Toomey)

R for Data Science (Dan Toomey). This e-book is written for data analysts who have a firm grip over advanced data analysis techniques. Some basic knowledge of the R language and some data science topics is also required. This book assumes that you have access to an R environment and are comfortable with the statistics involved.

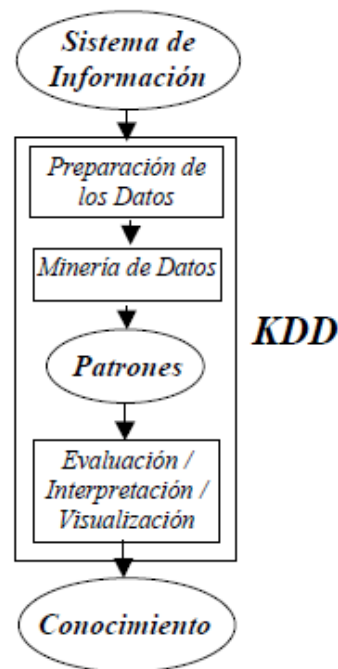


Statistique textuelle (Lebart y Salem, 1994)

Statistique textuelle (Lebart y Salem, 1994). Manual francés clásico de estudio de la estadística lexical.

1 Modulo 1 del curso “Empleo de R *Text Mining Solution* en el análisis avanzado de documentos de texto”; cursos on line de la FGUMA (año académico 2016-2017); recuperado el 1 de febrero de 2019, disponible en <https://fguma.cv.uma.es/course/view.php?id=2113>

Ejemplo 1: Aplicación de minería de textos en bases de datos bibliográficas



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.

1 Hernández Orallo J, Juan Lizandra M, Minaya Collado N & Montserrat Aranda C. Extracción y Visualización de Conocimiento de Bases de Datos Médicas; recuperado el 1 de febrero de 2019, disponible en

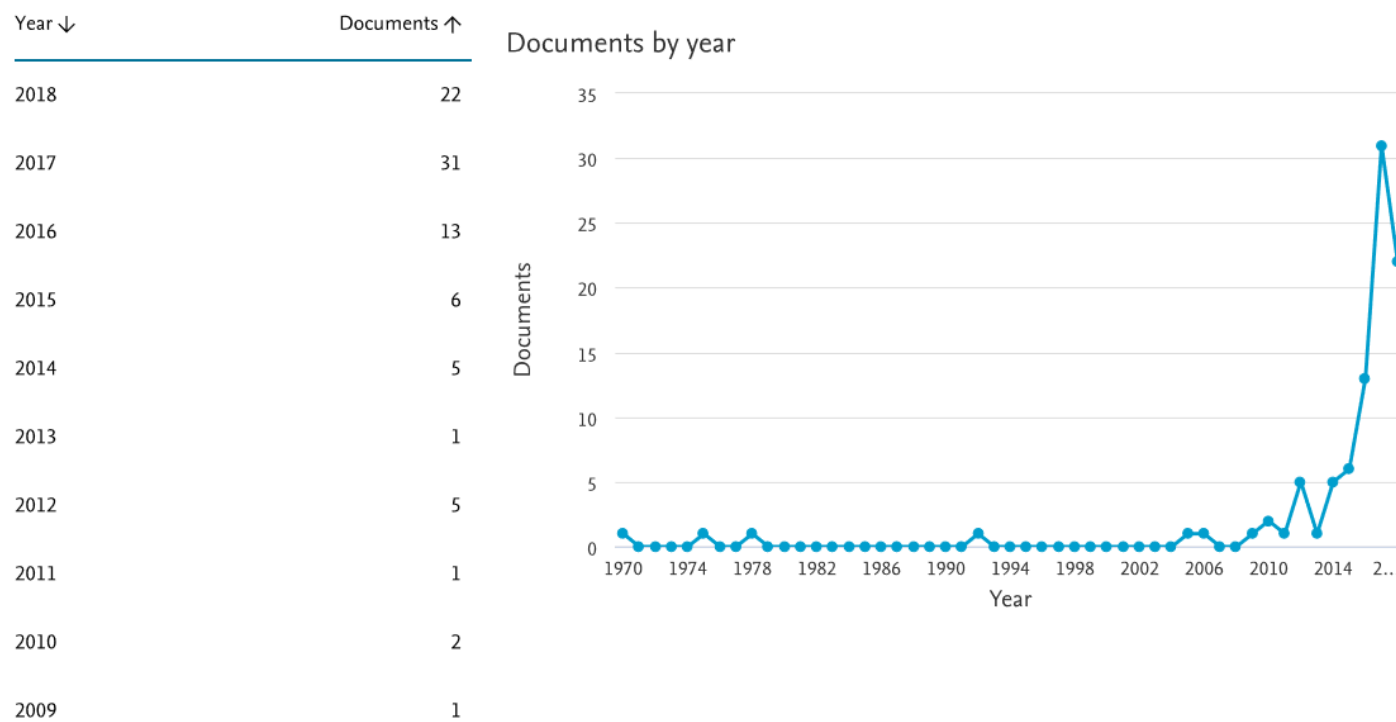
<https://www.researchgate.net/publication/242154806> Extraccion y Visualizacion de Conocimiento de Bases de Datos Medicas

Documentos en SCOPUS que contienen al menos uno de los términos *Talent Analytics*, *HR Analytics* o *People Analytics* en el título, resumen o palabras clave.

TITLE-ABS-KEY ("Talent Analytic*" OR "HR Analytic*" OR "People Analytic*")

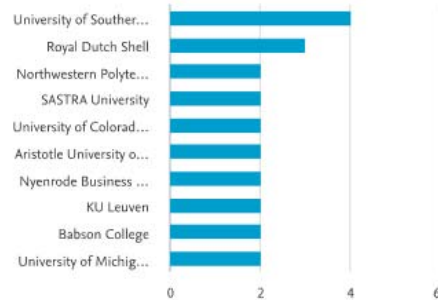
93 document results

Select year range to analyze: 1970 to 2018

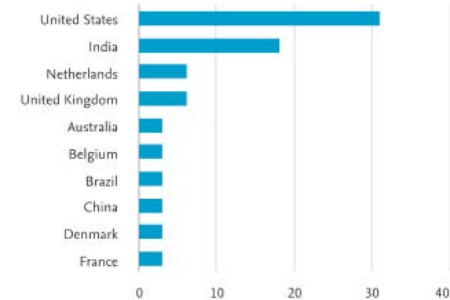


Características del corpus de documentos según la herramienta “Analyze search results” de SCOPUS.

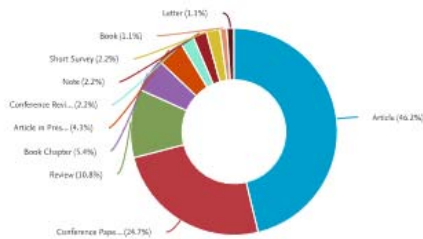
Documents by affiliation



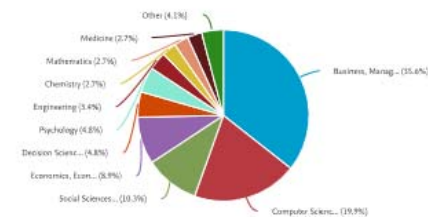
Documents by country/territory



Documents by type



Documents by subject area



Empleo del paquete Bibliometrix en la interface de RStudio

The screenshot displays the RStudio environment with the following components:

- Editor:** A script file named 'Untitled1.R' with a single line of code: `1`.
- Environment:** Shows 'Global Environment' with the message 'Environment is empty'.
- Files:** A table listing installed and available packages. The 'User Library' section is expanded to show the following packages:

Name	Description	Version
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-5
<input type="checkbox"/> acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
<input type="checkbox"/> assertthat	Easy Pre and Post Assertions	0.2.0
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.2
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> BH	Boost C++ Header Files	1.66.0-1
<input checked="" type="checkbox"/> bibliometrix	An R-Tool for Comprehensive Science Mapping Analysis	2.0.2
<input type="checkbox"/> bindr	Parametrized Active Bindings	0.1.1
<input type="checkbox"/> bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2
<input type="checkbox"/> ca	Simple, Multiple and Joint Correspondence Analysis	0.71
<input type="checkbox"/> car	Companion to Applied Regression	3.0-2
<input type="checkbox"/> carData	Companion to Applied Regression Data Sets	3.0-2
<input type="checkbox"/> cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0
<input type="checkbox"/> checkmate	Fast and Versatile Argument Checks	1.8.5
<input type="checkbox"/> cli	Helpers for Developing Command Line Interfaces	1.0.1
<input type="checkbox"/> colorspace	Color Space Manipulation	1.3-2
<input checked="" type="checkbox"/> colourpicker	A Colour Picker Tool for Shiny and for Selecting Colours in Plots	1.0
<input type="checkbox"/> cowplot	Streamlined Plot Theme and Plot Annotations for 'ggplot2'	0.9.3

- Console:** Shows the execution of the following R code:

```
library(bibliometrix); #bibliometrix is not a library?  
To cite bibliometrix in publications, please use:  
  
Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11(4), pp 959-975, Elsevier.  
  
http://www.bibliometrix.org  
  
To start with the shiny web-interface, please digit:  
biblioshiny()  
> biblioshiny()  
Listening on http://127.0.0.1:5869  
.
```

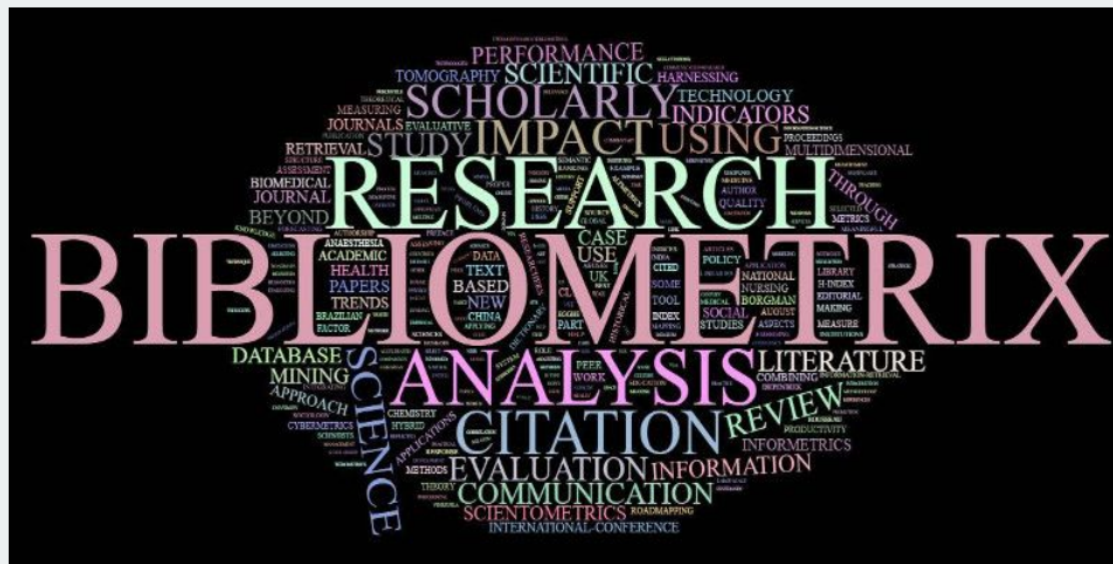
Análisis del corpus de documentos con la interface-web Biblioshiny

biblioshiny: The shiny app for bibliometrix R-package Welcome Load Filter Descriptive Analysis ▾ Conceptual Structure ▾ Intellectual Structure ▾ Social Structure ▾ Quit ▾

biblioshiny: The shiny app for bibliometrix

bibliometrix: An R-tool for comprehensive science mapping analysis

Aria, M., & Cuccurullo, C. (2017). *bibliometrix: An R-tool for comprehensive science mapping analysis*. *Journal of Informetrics*, 11(4), 959-975.



Análisis del corpus de documentos con la interface-web Biblioshiny

biblioshiny: The shiny app for bibliometrix R-package Welcome Load Filter Descriptive Analysis ▾ Conceptual Structure ▾ Intellectual Structure ▾ Social Structure ▾ Quit ▾

Database

Scopus ▾

Choose a file

Browse... scopus.bib.bib

Upload complete

Here accept single .txt/.bib/.xlsx/.RData files, or multiple .txt/.bib files compressed in a single .zip archive.

Conversion results

Number of Documents 93

Save as:

No, thanks! ▾

Show 50 rows
Copy
PDF
Print

Search:

DOI	AU	TI	SO	JI	DT	DT1	DE	ID	AB	C1
10.1016/J.IJINFORMGT.2018.08.002	TURSUNBAYEVA A;DILAUR S;PAGLIARI C	PEOPLE ANALYTICS—A SCOPING REVIEW OF CONCEPTUAL BOUNDARIES AND VALUE PROPOSITIONS	INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT	INT J INF MANAGE	REVIEW	ARTICLE	ADMINISTRATIVE DATA ANALYTICS; BUSINESS ANALYTICS; BUSINESS INFORMATICS; HR ANALYTICS; HRIS; HUMAN RESOURCE MANAGEMENT; PEOPLE ANALYTICS; TALEN	COMMERCE; HUMAN RESOURCE MANAGEMENT; PUBLIC ADDRESS SYSTEMS; BUSINESS ANALYTICS; BUSINESS INFORMATICS; DATA ANALYTICS; HR ANALYTICS; HRIS; PEOP	THIS MIXED-METHOD 'SCOPING REVIEW' MAPPED THE EMERGENCE OF THE TERM PEOPLE ANALYTICS (PA), THE VALUE PROPOSITIONS OFFERED BY VENDORS OF PA TOOLS AND S	EHEALTH RE: GROUP; U: INSTITUT POPULATION SCIENCES INFORMATI UNIVERSIT EDINBURGH, PLACE, EDINBURG 9AG,
10.1109/ASONAM.2018.8508512	MAURYA A;AKOGLU L;KRISHNAN R;BAYD	A LENS INTO EMPLOYEE PEER REVIEWS VIA SENTIMENT-ASPECT MODELING	PROCEEDINGS OF THE 2018 IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING, ASONAM 2018	PROC. IEEE/ACM INT. CONF. ADV. SOC. NETWORKS ANAL. MINING, ASONAM	CONFERENCE PAPER	CONFERENCE		REGRESSION ANALYSIS; VECTOR SPACES; ASPECT MODEL; LOGLINEAR MODEL; PEER REVIEW; REVIEW DATUM; SENTIMENT SCORES; VECTOR REPRESENTATIONS; VECTOR-	GIVEN A CORPUS OF EMPLOYEE PEER REVIEWS FROM A LARGE CORPORATION WHERE EACH REVIEW IS STRUCTURED INTO PROS AND CONS, WHAT ARE THE PREVALENT TRAITS THA	HEINZ COLL INFORMATION AND PUBLIC CARNEGIE M UNIVERSITY, STATES; UBER FRANCISCO, C STATE
10.1109/SEAA.2018.00042	PAPOUTOGLOU M;KAPITSAKI GM;MITTAS N	LINKING PERSONALITY TRAITS AND INTERPERSONAL SKILLS TO GAMIFICATION AWARDS	PROCEEDINGS - 44TH EUROMICRO CONFERENCE ON SOFTWARE ENGINEERING AND ADVANCED APPLICATIONS, SEAA 2018	PROC. - EUROMICRO CONF. SOFTW. ENG. ADV. APPL., SEAA	CONFERENCE PAPER	CONFERENCE	COMPETENCE MINING; PEOPLE ANALYTICS; PERSONALITY TRAITS; SOFTWARE REPOSITORIES	HUMAN ENGINEERING, COMPETENCE MINING; GENERALIZED LINEAR MODEL; INTERPERSONAL SKILLS; NON-TECHNICAL SKILLS; PEOPLE ANALYTICS; PERSONALITY TRAITS;	SOFTWARE REPOSITORIES AND QUESTION & ANSWER SITES CONSTITUTE PROMISING AND EVER-INCREASING SOURCES OF INFORMATION FOR SOFTWARE ANALYTICS AND FOR THE I	DEPARTME INFORMA ARISTOTLE UN OF THESSAL THESSALONIKI DEPARTME COMPUTER S UNIVERSITY OI NICOSI

Conceptual Structure, Thematic Map

Apply!

TM Parameters:

Field
Author's Keywords ▾

Min. word frequency
1 ▾

Thematic Map

Map

Table

Show 50 rows

Copy

Excel

PDF

Print

Search:

Occurrences	Words	Cluster	Cluster_Label
All	All	All	JOB <input type="text"/>
1	DOMINANCE ANALYSIS	2	JOB SATISFACTION
1	EMPLOYEE REVIEW DATA ANALYSIS	2	JOB SATISFACTION
1	INDUCTIVE REASONING	2	JOB SATISFACTION
2	JOB SATISFACTION	2	JOB SATISFACTION
1	LDA TOPIC MODELING	2	JOB SATISFACTION
1	TEXT MINING	2	JOB SATISFACTION
1	PERSONNEL RECORDS	2	JOB SATISFACTION
1	QUITS	2	JOB SATISFACTION
1	RETENTION	2	JOB SATISFACTION
1	SURVEY DATA	2	JOB SATISFACTION

Showing 1 to 10 of 10 entries (filtered from 228 total entries)

Previous

1

Next

Thematic Map, diagrama estratégico de temas

biblioshiny: The shiny app for bibliometrix R-package

Welcome

Load

Filter

Descriptive Analysis ▾

Conceptual Structure ▾

Intellectual Structure ▾

Social Structure ▾

Quit ▾

Apply!

TM Parameters:

Field

Author's Keywords

Min. word frequency

1

Words

Cluster

Cluster_Label

text

All

All

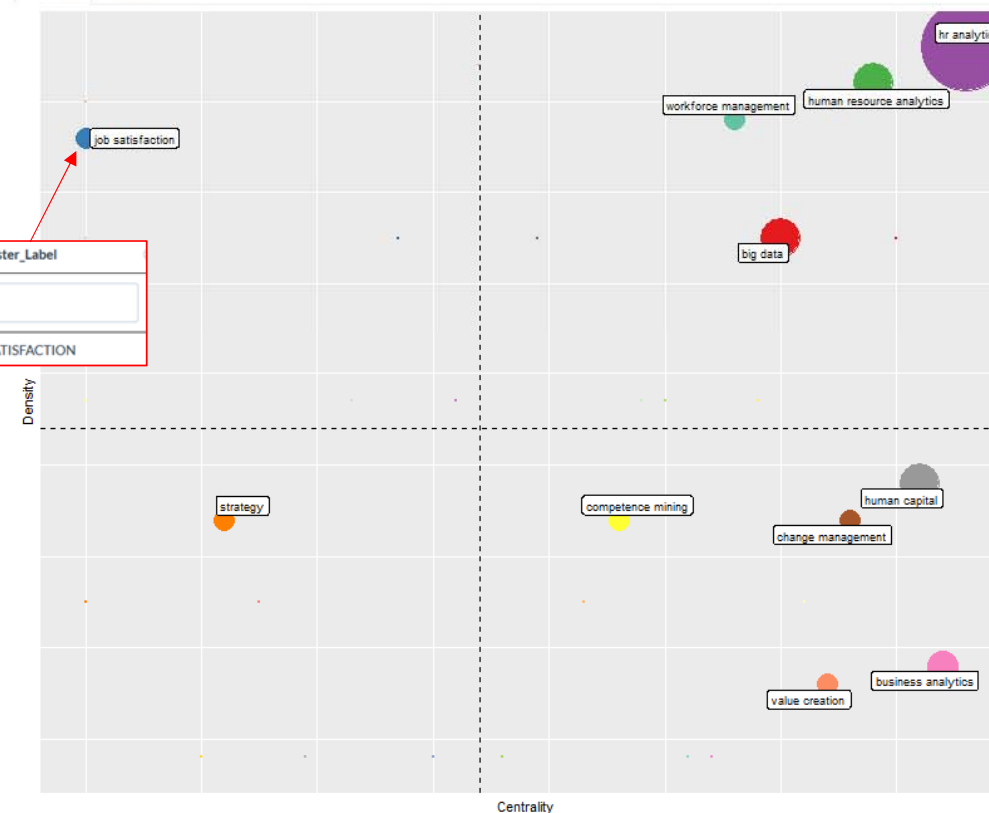
TEXT MINING

2

JOB SATISFACTION

Thematic Map

Map Table



Documento con el término “Text Mining” que forma parte del corpus recuperado de SCOPUS

Document details

< Back to results | 1 of 1

BibTeX export ▾ Download Print E-mail Save to PDF Add to List More... >



ICIS 2017: Transforming Society with Digital Innovation
2018

38th International Conference on Information Systems: Transforming Society with Digital Innovation, ICIS
2017; CoexSeoul; South Korea; 10 December 2017 through 13 December 2017; Code 133354

A Study on Job Satisfaction Factors in Retention and Turnover Groups using
Dominance Analysis and LDA Topic Modeling with Employee Reviews on
Glassdoor.com (Conference Paper)

Lee, J.^a ✉, Kang, J.^b ✉

^aSchool of Integrated Technology, Yonsei University, Incheon, South Korea

^bDepartment of e-Business, Ajou University, Suwon, South Korea

Author keywords

Dominance Analysis Employee Review Data Analysis Inductive Reasoning Job Satisfaction LDA Topic Modeling
Text mining

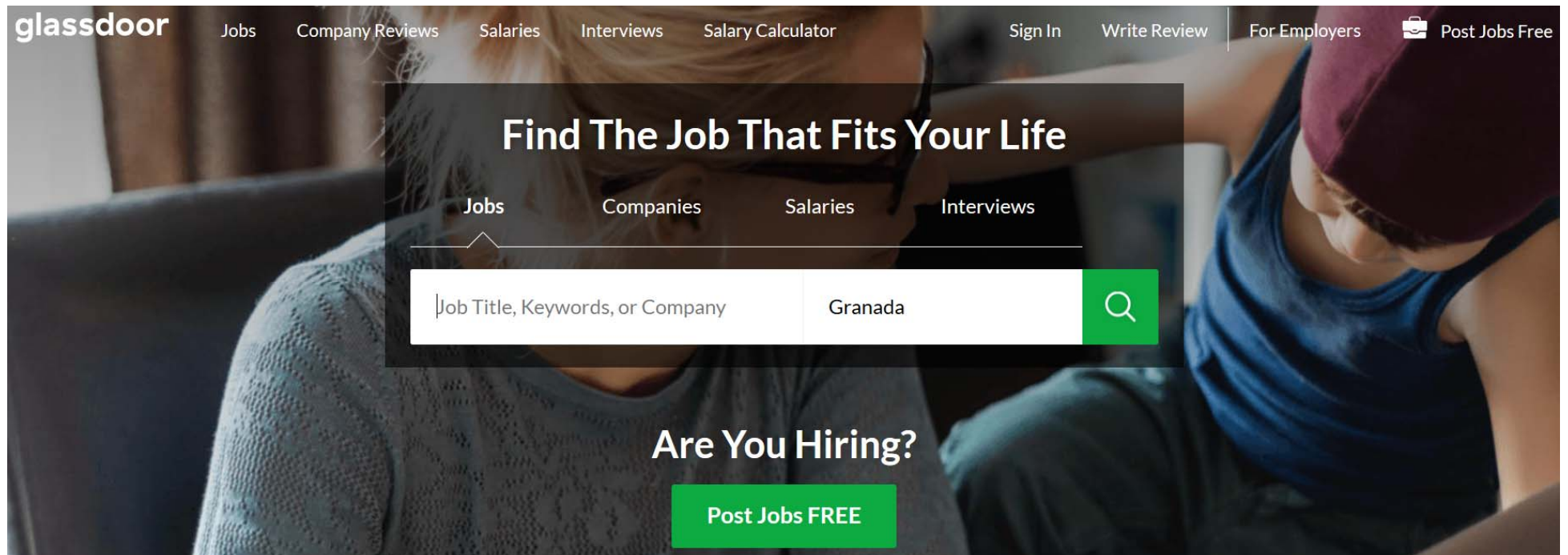
Indexed keywords

Engineering controlled terms: Big data Data handling Factor analysis Human resource management
Information analysis Job satisfaction

Engineering uncontrolled terms: Dominance analysis Inductive reasoning Review datum Text mining Topic Modeling

Engineering main heading: Data mining

Glassdoor.com: portal web que reúne comentarios de empleados



Estudio de la satisfacción laboral según los comentarios de empleados en Glassdoor.com

A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor.com

Short Paper

Jongseo Lee

School of Integrated Technology,
Yonsei University
Incheon, Korea
semanticweb7@yonsei.ac.kr

Juyoung Kang

Department of e-Business,
Ajou University
Suwon, Korea
jykang@ajou.ac.kr

Abstract

HR analytics is an important area for the application of big data analysis techniques, and the organizational insight that it provides enables effective management of employees. In this paper, we analyze employee review data posted on a representative third-party employee review website. We identify the relative importance of factors affecting job satisfaction and then extract topic differences after classifying employees according to retention and turnover. First, LDA Topic Modeling by adopting n -grams is performed on unstructured text data to analyze employee review data. Second, a dominance analysis is conducted to examine the relative importance of job factors. We found that the "Culture and Values" and "Senior Management" factors have the highest influence on both retention and turnover. Our model follows a novel approach in applying the analysis of reviews and text mining to the HR domain and will be of practical relevance for enhancing employee retention.

Keywords: Employee Review Data Analysis, Job Satisfaction, Text mining, Dominance Analysis, LDA Topic Modeling, Inductive Reasoning

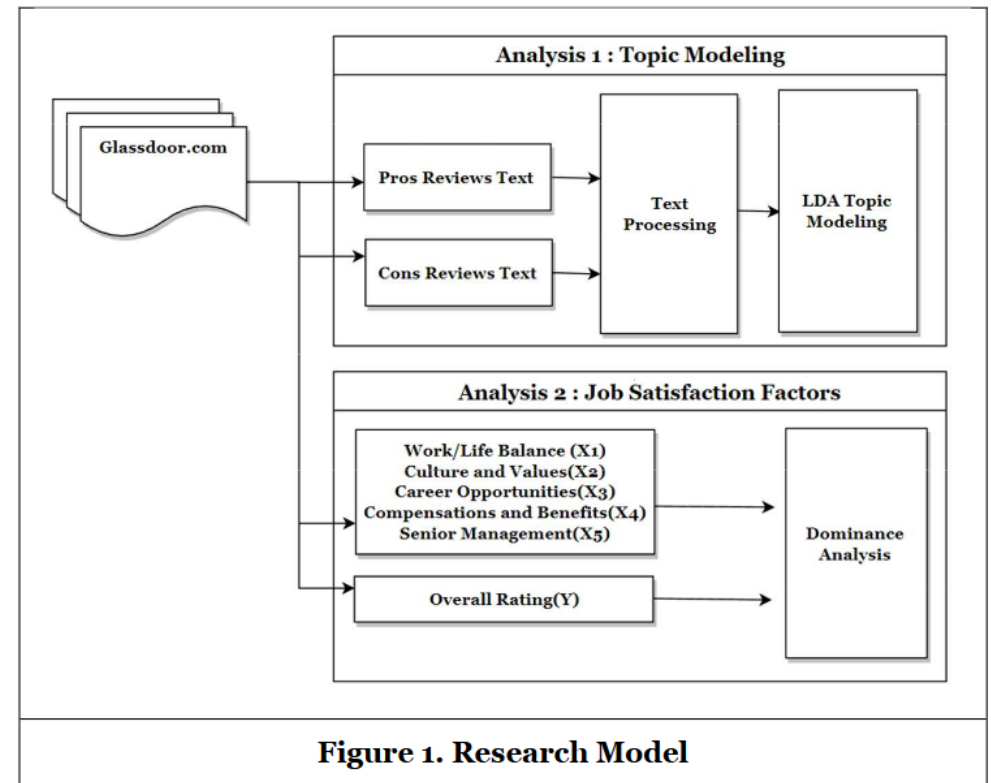
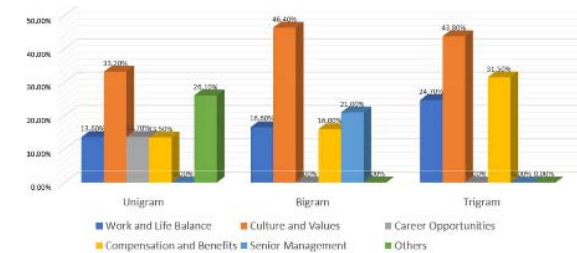


Figure 1. Research Model

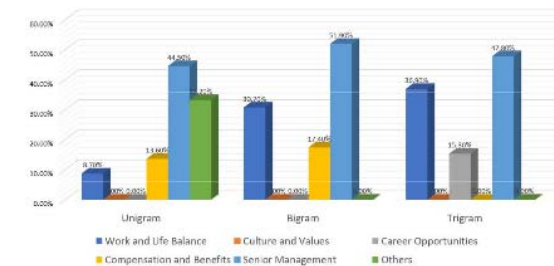
“LDA topic modeling”, una técnica de minería de textos que extrae temas

Table 2. Extracted Relevant Keywords for Employee Retention and Turnover

Group	Unigram	Bigram	Trigram
Retention Employees With Positive Ratings (Pros)	work, benefits, career, growth, flexible, culture, opportunities, free, fun, pay, friendly, good, great, time, health, insurance, vacation, advancement, ...	work life, life balance, great benefits, work environment, flexible work, tuition reimbursement, great people, friendly people, balance good, great pay, part time,	work life balance, great work environment, life balance good, good benefits salaries, great company culture, friendly work environment, flexible work schedule, ...
Turnover Employees With Negative Ratings (Cons)	leadership, management, culture, people, time, hours, poor, life, sales, low, lack, benefits, balance, pay, layoffs, terrible, managers, growth, training, ...	senior management, upper management, long term, work life, life balance, low pay, long hours, poor management, minimum wage, senior leadership, ...	work life balance, poor work life, high turnover rate, limited growth opportunities, stressful work environment, long term employees, management doesn't care, ...

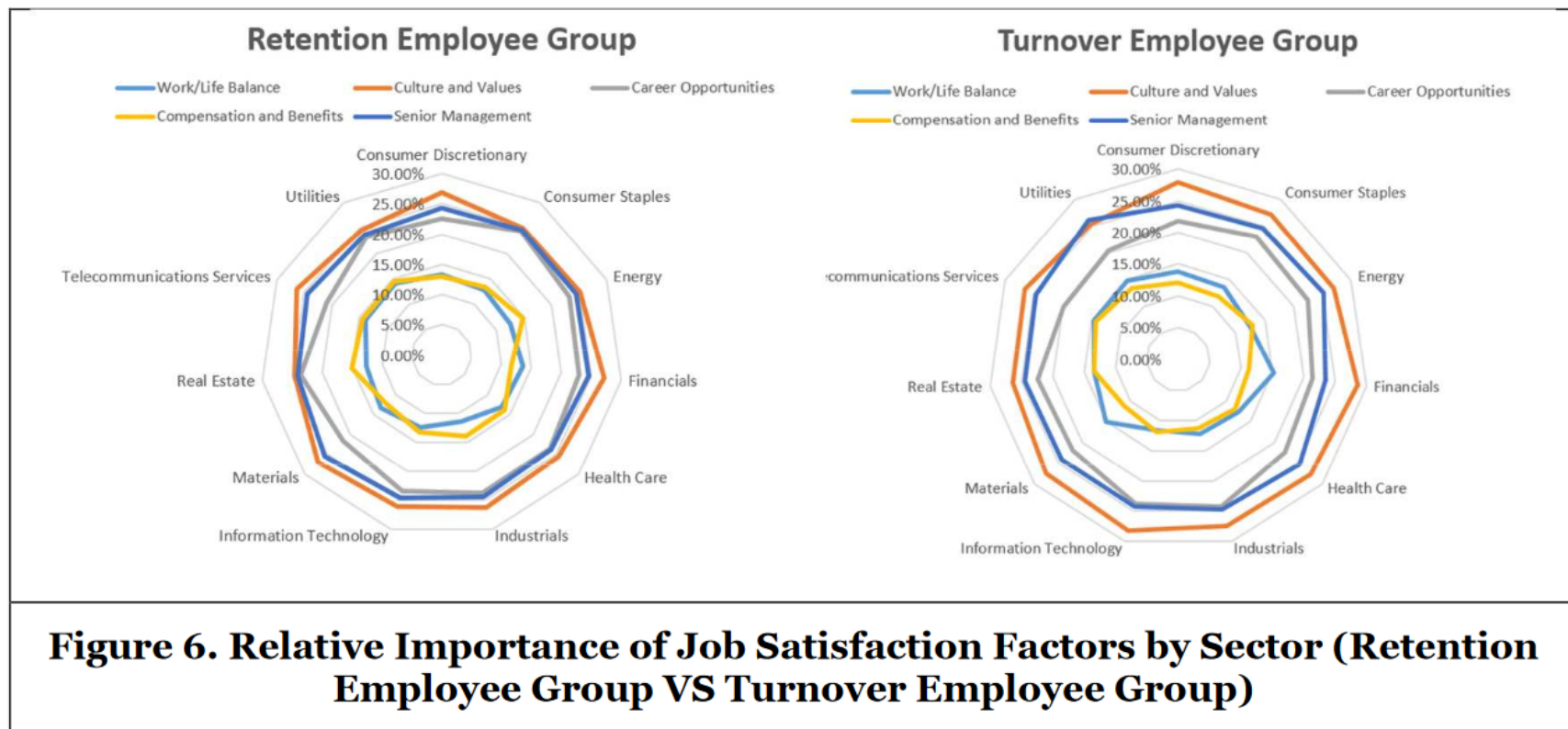


Topic Modeling Composition of Pros for Retention with Positive Ratings



Topic Modeling Composition of Cons for Turnover with Negative Rating

Importancia relativa de los factores de satisfacción laboral por sector (grupo de empleados: retención /rotación)



Toma de decisiones -> Para elaborar una estrategia para la retención de empleados basada en el análisis de datos

Las organizaciones deben enfocarse en crear una cultura corporativa que satisfaga las necesidades de los empleados, alentándolos a identificarse y adaptarse fácilmente a ella.

El factor de la Alta Dirección también es relativamente importante en ambos grupos, lo que indica que la relación de los empleados con la administración es crucial. La alta gerencia debe garantizar la comunicación con frecuencia con sus empleados de una manera accesible e inclusiva, a través de diversas vías, para que los empleados se sientan lo suficientemente valorados y seguros para expresar sus quejas y hacer contribuciones positivas.

Desarrollar un modelo predictivo para detectar tendencias que indiquen la posible rotación futura de los empleados en una etapa preventiva. Este modelo predictivo ayudaría a la administración a reaccionar de manera más oportuna y proactiva a la pérdida inminente de recursos humanos clave.

Ejemplo 2: Aplicación de minería de textos en análisis de las 215 respuestas de encuestas abiertas



Con el objeto de mejorar futuros MOOC y de continuar avanzando en una cultura de la innovación te pedimos valores de 1 a 10 los siguientes aspectos del que has realizado:

Vamos a analizar las respuestas al apartado 1 de la encuesta: “Nivel de consecución de los siguientes Objetivos del MOOC”

Con el objeto de mejorar futuros MOOC y de continuar avanzando en una cultura de la innovación te pedimos valores de 1 a 10 los siguientes aspectos del que has realizado:

1. Nivel de consecución de los siguientes Objetivos del MOOC

- Sensibilizar y motivar a los y las participantes para extender una cultura de la innovación en la Junta de Andalucía.
- Identificar las claves de la innovación en el sector público.
- Facilitar el intercambio, la interacción y el debate colectivo entre participantes en el MOOC, creando dinámicas de trabajo colaborativo.
- Familiarizarse con el uso de herramientas tecnológicas de información y comunicación
- Conocer experiencias de innovación en el sector público.
- Facilitar recursos y herramientas que permitan definir e implementar retos y proyectos de innovación.

Comentarios:

Se empleará el paquete RcmdrPlugin.temis

← → ↻ 🏠

🔒 <https://rtemis.hypotheses.org> 67% ⋮ 🛡️ ☆

R.TeMiS

Une approche intégrée et libre de l'analyse de données textuelles

[À propos](#) [Téléchargement & Installation](#) [Utilisation](#) [Exportation de corpus depuis Factiva](#) [Publications](#) 🔍

À propos

R.TeMiS [R Text Mining Solution] est un environnement graphique de travail sous R permettant de créer, manipuler et analyser des corpus de textes. Il a été conçu pour limiter les effets de « boîte noire », auxquels sont souvent confrontés les utilisateurs de logiciels de statistique lexicale, et favoriser la réflexivité dans l'usage sociologique des données textuelles.

L'architecture statistique de l'environnement **R.TeMiS** est fournie par le paquet `tm` développé par Ingo Feinerer (Feinerer, 2008; 2011; Feinerer, Hornik & Meyer, 2008). Celui-ci a été complété par d'autres paquets classiques de R comme `ca` pour la représentation des analyses factorielles des correspondances (Nenadic & Greenacre, 2007). Enfin des paquets spécifiques ont été développés pour faciliter l'usage de **R.TeMiS** dans le domaine des études sur les médias, par exemple pour la gestion des corpus constitués depuis la base de données d'articles de presse Factiva.

Afin de faciliter l'usage de **R.TeMiS** aux néo-utilisateurs de R, le développement d'un environnement graphique a été privilégié. Celui-ci se présente donc comme un menu de R Commander (Fox, 2005).

R.TeMiS est développé par [Milan Bouchet-Valat](#) (Ined) et [Gilles Bastin](#) (Sciences Po Grenoble, Pacte).

Nouvelles de R.TeMiS

- [Nouvelle version 0.7.10 22/06/2018](#)
- [Nouvelle version 0.7.9 11/07/2017](#)
- [Nouveau tutoriel pas à pas 19/05/2017](#)
- [Nouvelle version 0.7.8 02/11/2016](#)
- [Nouvelle version 0.7.7 08/07/2016](#)

Administration

- [Connexion](#)
- [Flux RSS des articles](#)
- [RSS des commentaires](#)
- [Site de WordPress-FR](#)

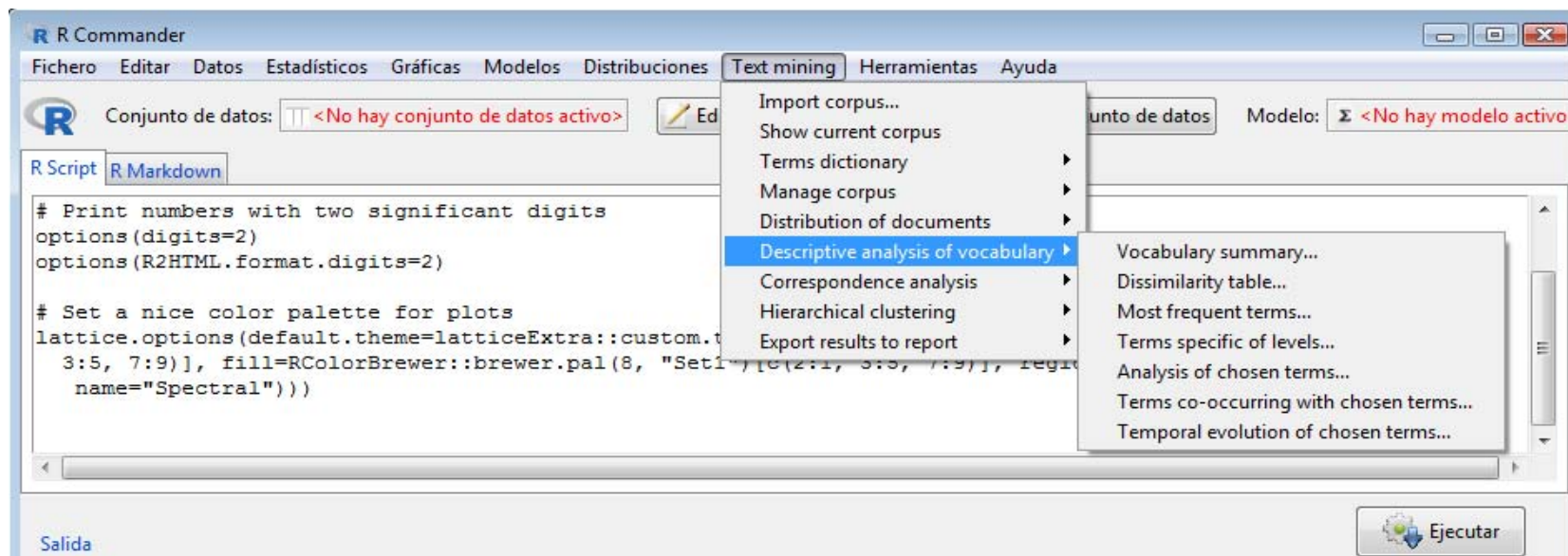
R Text Mining Solution (R.TeMiS)

Bouchet-Valat, M., & Bastin, G. (2013). RcmdrPlugin.temis, a Graphical Integrated Text Mining Solution in R. *The R Journal* , 188-196.

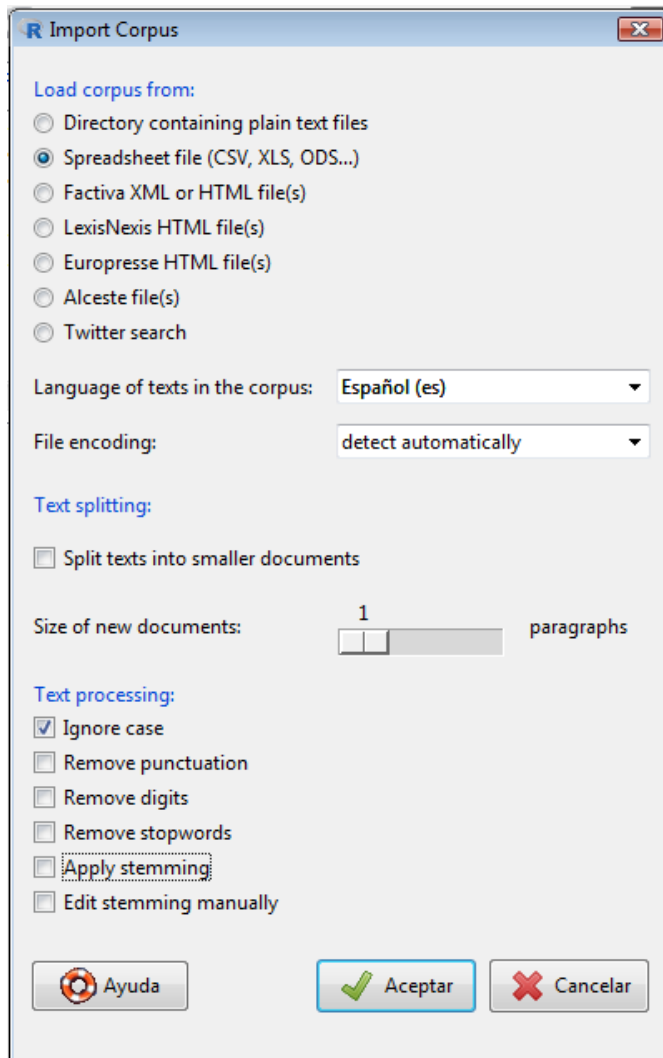
Bouchet-Valat, M. (2016). *Package RcmdrPlugin.temis*.

<https://cran.r-project.org/web/packages/RcmdrPlugin.temis/>

Garnier, B. (2014). *R.TeMiS. Une approche intégrée et libre de l'analyse de données textuelles*. <http://rtemis.hypotheses.org/>

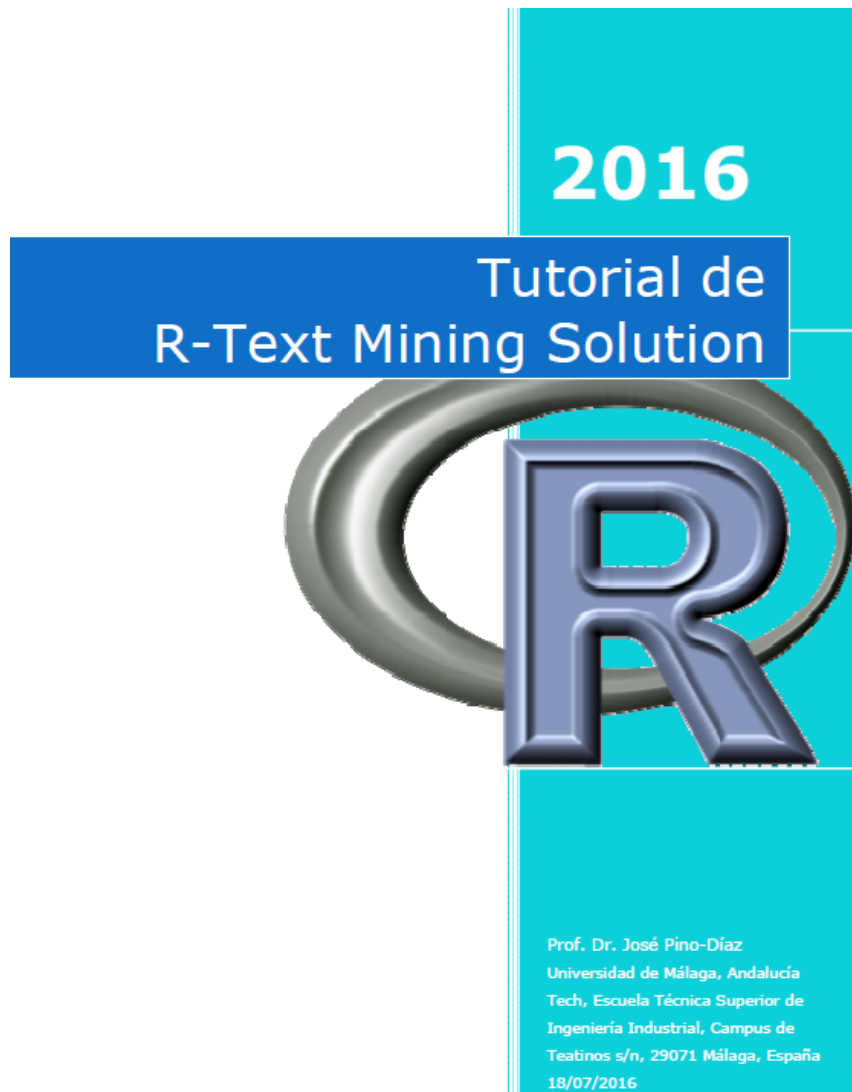


R Text Mining Solution (R.TeMiS)



1. Visualización del corpus activo y de los diccionarios de términos.
2. Gestión y distribución del corpus.
 1. Creación y análisis de subcorpus por términos o por variables
 2. Creación de tablas y gráficos de distribución.
3. Análisis descriptivo del léxico.
 1. Resumen cuantitativo del vocabulario de términos.
 2. Tabla de disimilaridad.
 3. Términos más frecuentes.
 4. Términos específicos por modalidades de la variable.
 5. Análisis de términos concretos.
 6. Términos que coocurren con otros concretos.
 7. Evolución temporal de términos concretos.
4. Análisis factorial de correspondencias.
 1. AFC de la matriz documentos-términos sin agregar ninguna variable.
 2. AFC de la tabla lexical completa agregando variables.
5. Clasificación ascendente jerárquica.

Tutorial R.TeMiS (Pino, 2016)

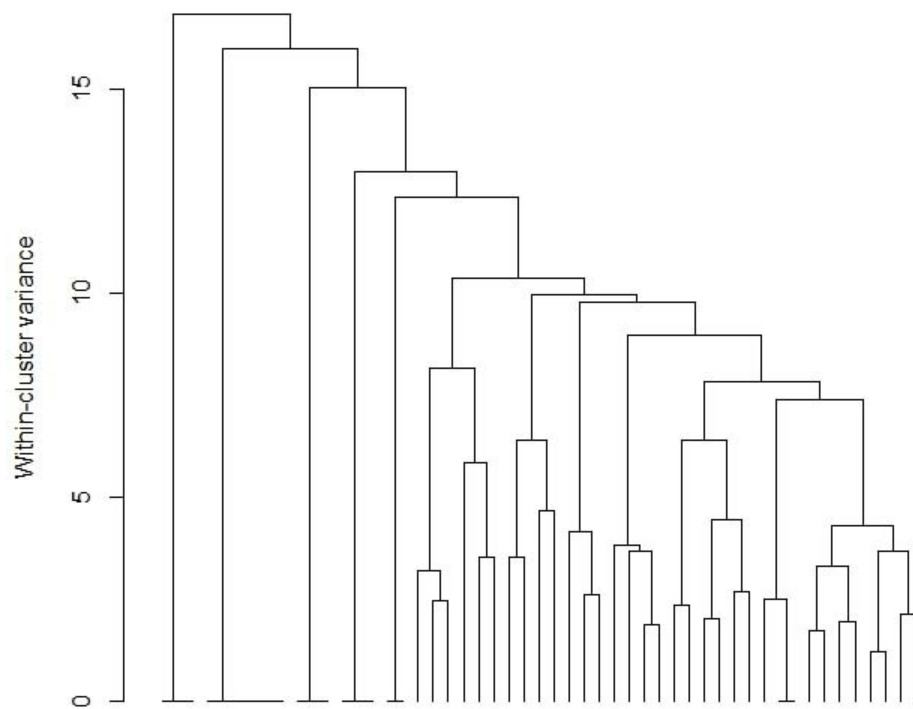


Contenido

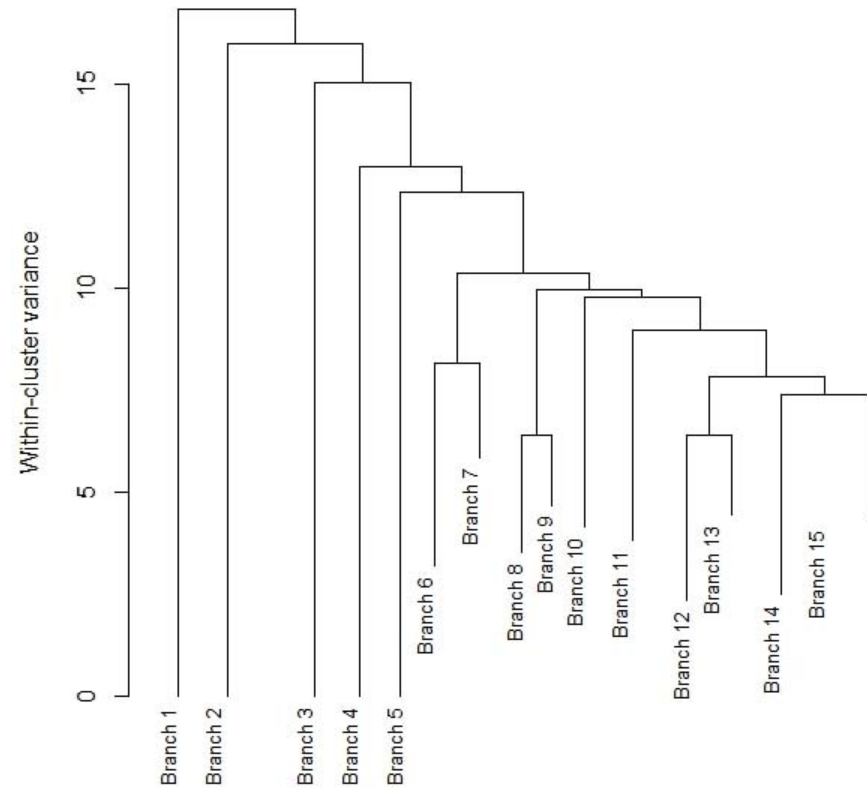
Introducción	4
Comenzando a trabajar con R.TeMiS	5
Importar Corpus	8
Visualización del corpus activo y de los diccionarios de términos	13
Visualizar el corpus de documentos.....	13
Visualizar el diccionario de términos.....	14
Gestión y distribución del corpus.....	15
Análisis descriptivo del léxico.....	21
Resumen cuantitativo del vocabulario de términos.....	22
Tabla de disimilaridad.....	23
Términos más frecuentes.....	24
Términos específicos por modalidades de la variable.....	25
Análisis de términos concretos.....	27
Términos que coocurren con otros concretos.....	29
Evolución temporal de términos concretos.....	29
Análisis de correspondencias aplicado a un corpus de documentos	30
Concepto de similaridad, disimilaridad, distancia y proximidad entre documentos.....	30
Análisis de correspondencias.....	30
Análisis de correspondencias con RTemis.....	31
Interpretación del diagrama de correspondencias	34
Procedimiento de análisis de correspondencias con RTemis.....	35
I) AFC de la matriz documentos-términos sin agregar ninguna variable ...	35
II) AFC de la tabla lexical completa agregando variables (<i>full document- term matrix by variables</i>).....	40
Clasificación ascendente jerárquica aplicada a un corpus de documentos	45
Concepto de similaridad y distancia entre documentos.....	45
Concepto de clasificación jerárquica.....	45
Clasificación ascendente jerárquica.....	46
Clasificación ascendente jerárquica con RTemis.....	47

Análisis con R.TeMiS: Clasificación ascendente jerárquica

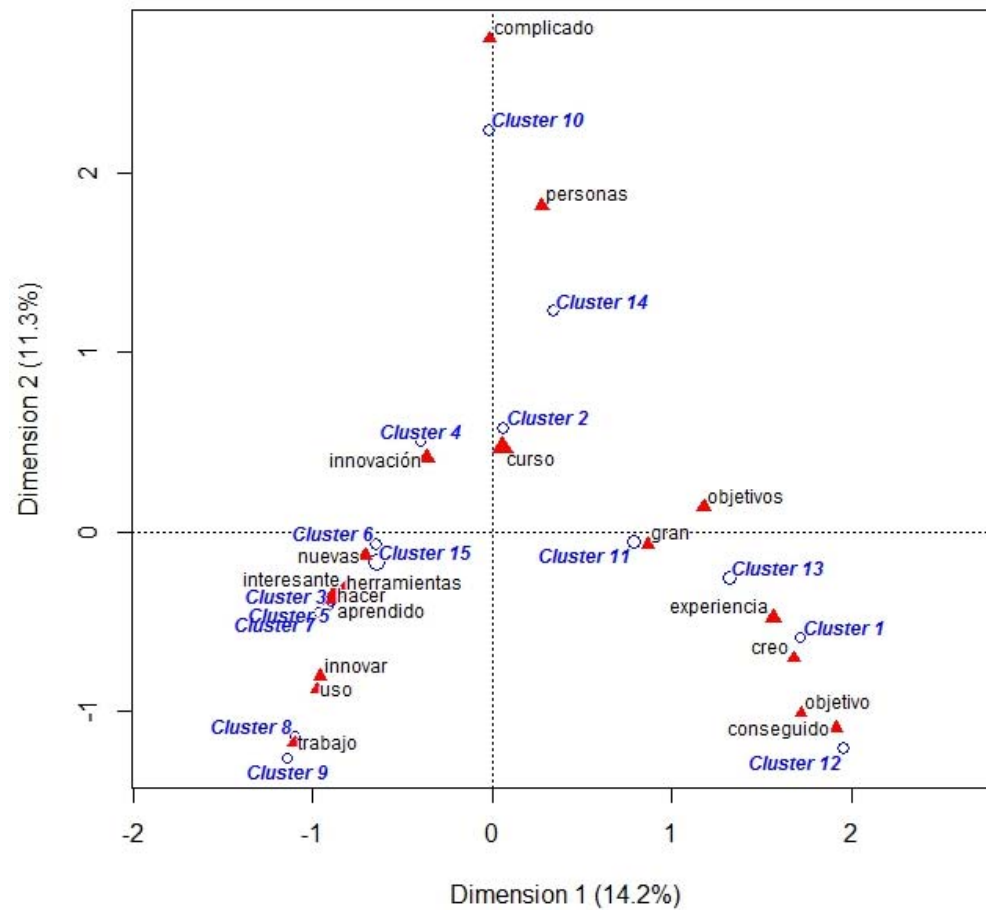
Full documents dendrogram



Clusters dendrogram



Análisis con R.TeMiS: Análisis factorial de correspondencias



Resultados

Comentarios al apartado 1 de la encuesta: "Nivel de consecución de los siguientes Objetivos del MOOC"			
Grupos	Términos significativos	A favor (+)	En contra (-)
Grupo 1	"experiencia"	Experiencia positiva. Me ha gustado la experiencia.	Excesiva burocracia.
Grupo 2	"curso"	Me ha entusiasmado. Fascinante. Me ha encantado. Curso motivador. Curso diferente. Curso completo.	Me ha decepcionado un poco la insignia. Me hubiera gustado tener un certificado del curso para poder aportarlo como mérito
Grupo 3	"herramientas"	Me ha aportado conocimientos, herramientas y recursos muy útiles. Me he iniciado en el uso de nuevas herramientas que desconocía.	Dificultad en el manejo de herramientas tecnológicas que facilitan el trabajo colaborativo
Grupo 4	"innovación"	Buena toma de contacto con el tema de la innovación. Se da a conocer la innovación en el trabajo del empleado público.	He encontrado a faltar experiencias de innovación en sistemas de gestión de RR.HH Parecía más un curso de redes sociales y autobombo que de innovación en sí
Grupo 5	"interesante"	Interesante por lo innovador en el método didáctico	
Grupo 6	"nuevas"	Abre nuevos horizontes. He hecho cosas nuevas. He practicado con nuevas tecnologías	
Grupo 7	"aprendido", "hacer"	He aprendido a hacer cosas nuevas. He conocido iniciativas muy interesantes de compañeros.	
Grupo 8	"innovar"	Se puede innovar en el trabajo. Innovar es importante.	Más empeño en twittear que en innovar.
Grupo 9	"puestos"		Impedimentos técnicos para realizar el curso en algunos puestos de trabajo. La realidad en los puestos de trabajo es otra a la planteada en el curso.
Grupo 10	"complicado"		Curso complicado para personas sin mínimos conocimientos en informática. La elevada participación y debate complica el resultado. En la Admon es complicado innovar desde la perspectiva del curso.
Grupo 11	"gran"	Gran seguimiento semanal Gran capacidad de motivación	
Grupo 12	"conseguido", "objetivo"	Objetivo conseguido	Decepcionada por la evaluación entre pares
Grupo 13	"objetivos"	Grado satisfactorio por conseguir los objetivos	
Grupo 14	"personas"	Curso muy motivador para personas con ganas de mejorar Curso muy activo donde personas novatas consiguen objetivos	No se expide certificado del curso para todas las personas
Grupo 15	"ideas"	Se comparten y se toman ideas.	

Toma de decisiones -> Para mejorar futuros MOOC y continuar avanzando en una cultura de la innovación en la Admon

La puntuación media de los seis objetivos del MOOC es de 8,3 sobre 10. Ahora bien, hay aspectos a mejorar. Según los comentarios al apartado 1 de la encuesta:

- a. Mejorar la insignia del curso.
- b. Expedir certificado para todos los participantes que lo superen, independientemente del tipo de Admon donde trabajen.
- c. Homogeneizar los grupos de participantes según el nivel de usuario en informática.
- d. Incluir y explicar experiencias de innovación en RR.HH.
- e. Facilitar la realización del curso, resolver los impedimentos técnicos en los centros.
- f. Acercar los planteamientos del curso a la realidad cotidiana de los centros.
- g. Optimizar el número de participantes para agilizar la participación y los debates.

¿Preguntas?... ¡Comentarios!...

Muchas Gracias

Prof. Dr. José Pino Díaz

Universidad de Málaga

Andalucía Tech

Escuela de Ingenierías Industriales

Departamento de Economía y Administración de Empresas,

Grupos de investigación Techné (UGR) e iArtHis-Lab (UMA)

Campus de Teatinos s/n, 29071 Málaga, España