

WISSENSCHAFTLICHES ARBEITEN

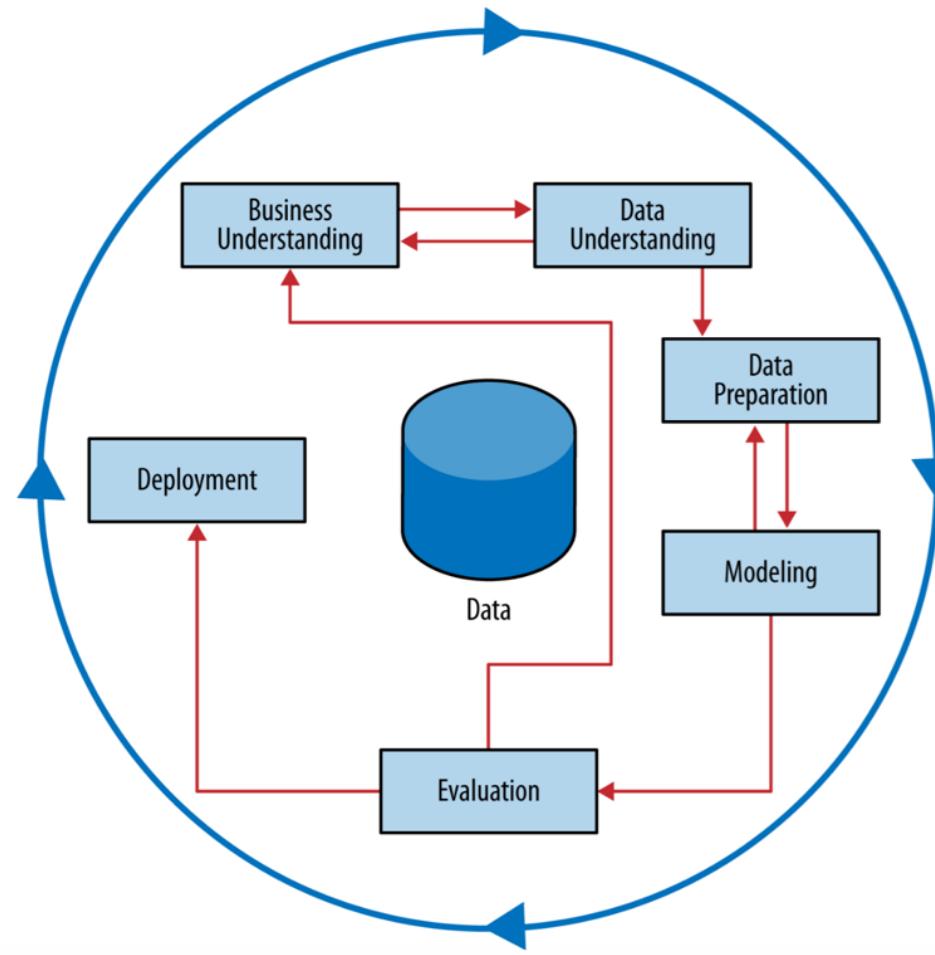
Predictive Analytics

Today's Agenda

- I. Short Repetition of Kick-off
- II. Science
- III. Reproducibility
- IV. Literature Review
- V. Q&A

SHORT REPETITION OF KICK-OFF

Cross-Industry Process for Data Mining (CRISP-DM)



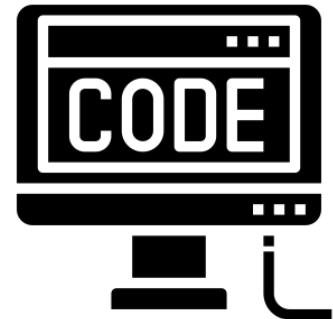
Source: Shearer (2000)

SHORT REPETITION OF KICK-OFF

Prüfungsleistungen (examinations)			
Art der Modulprüfung (type of modul examination): Modulprüfung			
Art der Prüfung (type of examination)	Umfang (extent)	Gewichtung (weighting)	
a) Hausarbeit mit Präsentation	ca. 15 Seiten	60.00 %	
b) Präsentation	20-30 Minuten	40.00 %	



+



R/Python Notebook

PDF Document



	Seminar Papers	Bachelor Thesis	Master Thesis
Topic/ Research Question	Given by supervisor	Find your own topic and RQ, based on existing literature	Find your own topic and RQ, based on real-life problem
Dataset	Given by supervisor	Select your own dataset (e.g., Kaggle)	Collect a “new” dataset
Structure	CRISP-DM	Like a scientific paper	Like a scientific paper
Literature Review	Business understanding, Modelling	+ Motivation & Research gap + Discussion	+ Implications for research and practice

SCIENCE

Generating new knowledge based on observations of the world through systematic and organised methods

Source: Recker (2012)

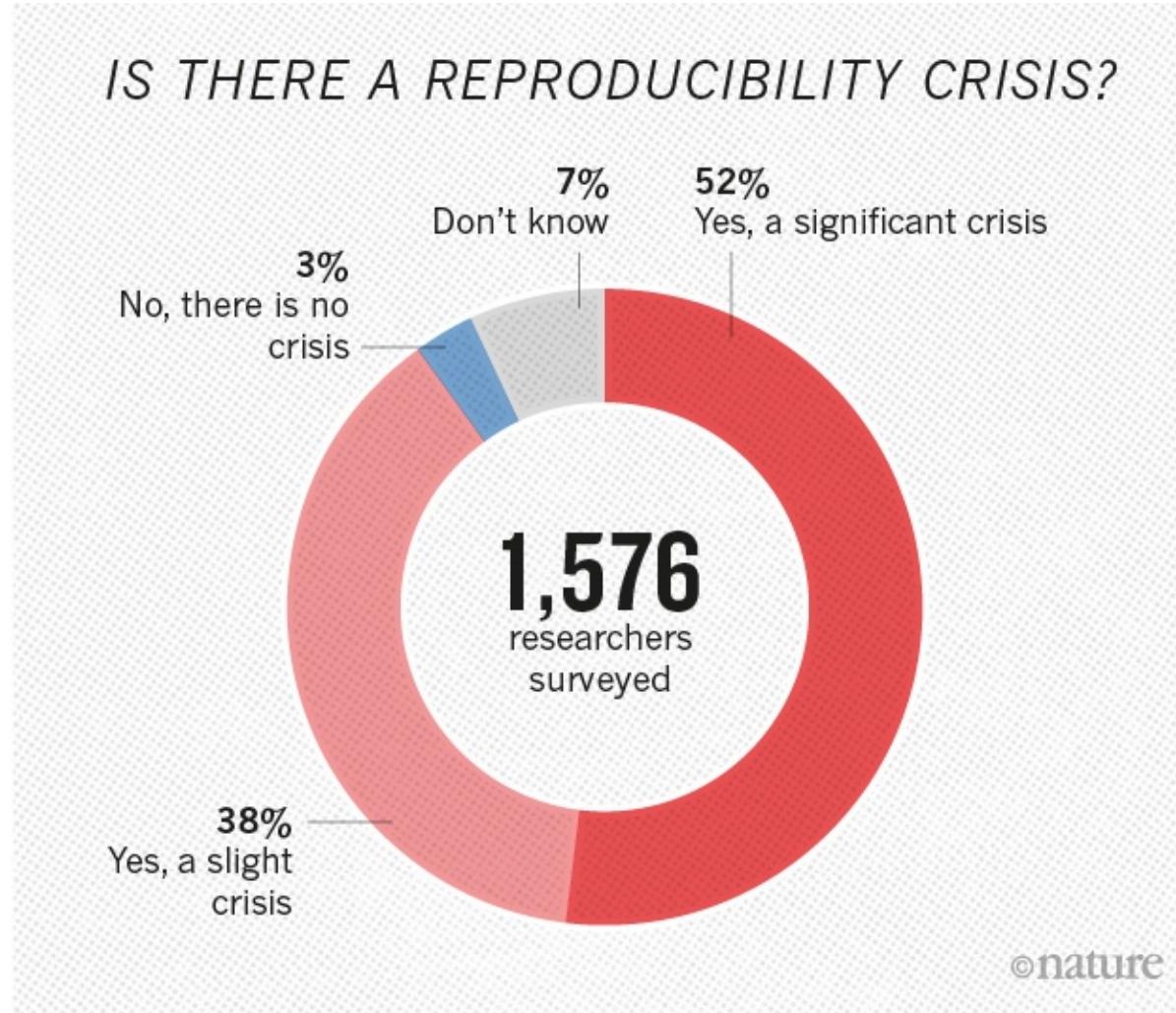
REPRODUCIBILITY

Reproducibility Crisis

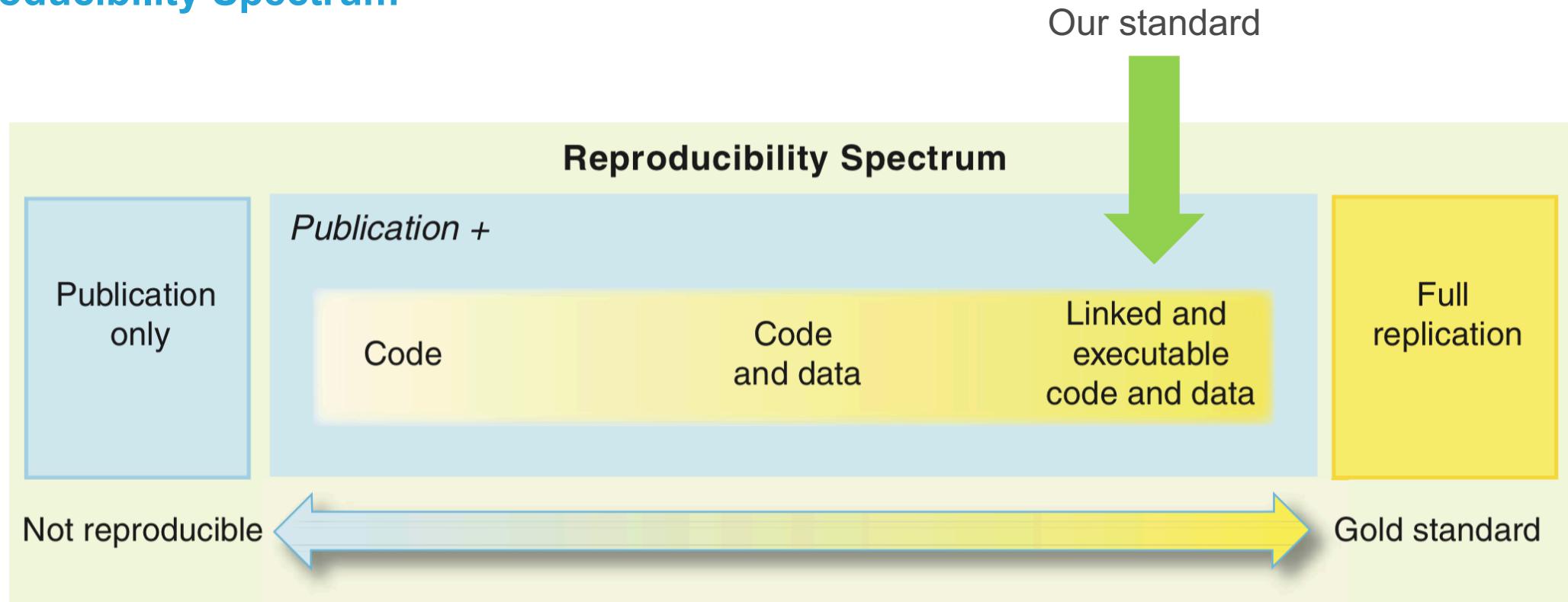


Source: <https://static.vecteezy.com>

Reproducibility Crisis



Reproducibility Spectrum



Source: Peng (2011)

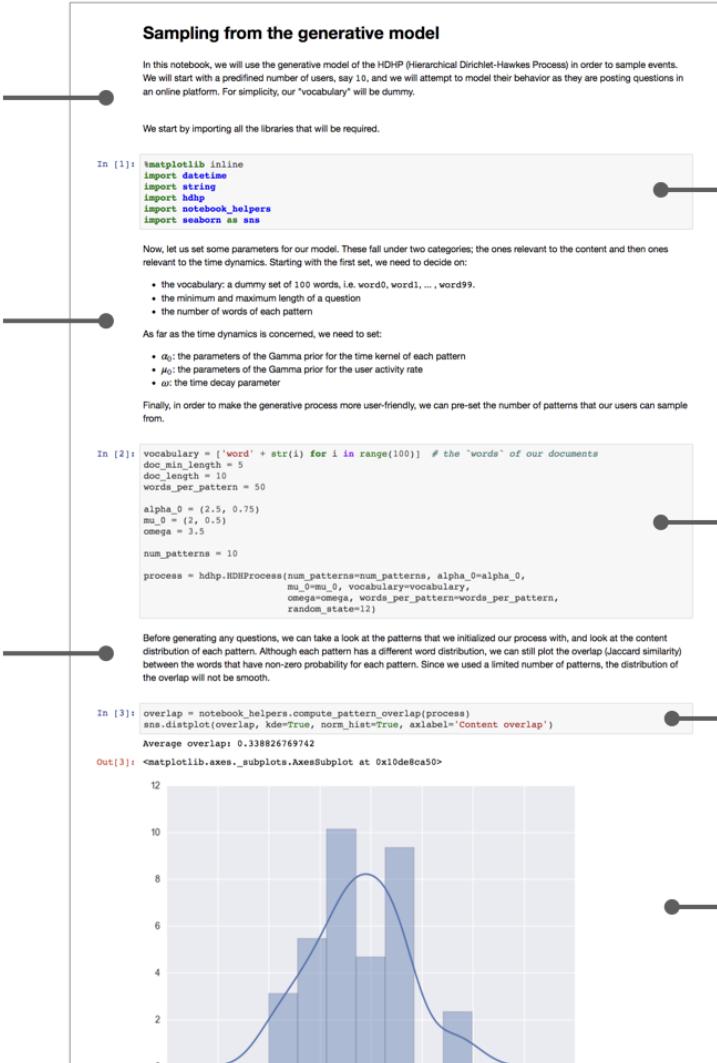
Jupyter Notebooks

Narrative Text

Notebook title and introduction

Description of model parameters

Description of need to profile data



Sampling from the generative model

In this notebook, we will use the generative model of the HDHP (Hierarchical Dirichlet-Hawkes Process) in order to sample events. We will start with a predefined number of users, say 10, and we will attempt to model their behavior as they are posting questions in an online platform. For simplicity, our "vocabulary" will be dummy.

We start by importing all the libraries that will be required.

```
In [1]: import matplotlib
import datetime
import string
import hdhp
import notebook_helpers
import seaborn as sns
```

Now, let us set some parameters for our model. These fall under two categories; the ones relevant to the content and then ones relevant to the time dynamics. Starting with the first set, we need to decide on:

- the vocabulary: a dummy set of 100 words, i.e. word0, word1, ..., word99,
- the minimum and maximum length of a question
- the number of words of each pattern

As far as the time dynamics is concerned, we need to set:

- α_0 : the parameters of the Gamma prior for the time kernel of each pattern
- μ_0 : the parameters of the Gamma prior for the user activity rate
- ω : the time decay parameter

Finally, in order to make the generative process more user-friendly, we can pre-set the number of patterns that our users can sample from:

```
In [2]: vocabulary = ['word' + str(i) for i in range(100)] # the 'words' of our documents
doc_min_length = 5
doc_length = 10
words_per_pattern = 50

alpha_0 = (2.5, 0.75)
mu_0 = (2, 0.5)
omega = 3.5

num_patterns = 10

process = hdhp.HDHPProcess(num_patterns=num_patterns, alpha_0=alpha_0,
                           mu_0=mu_0, vocabulary=vocabulary,
                           omega=omega, words_per_pattern=words_per_pattern,
                           random_state=12)
```

Before generating any questions, we can take a look at the patterns that we initialized our process with, and look at the content distribution of each pattern. Although each pattern has a different word distribution, we can still plot the overlap (Jaccard similarity) between the words that have non-zero probability for each pattern. Since we used a limited number of patterns, the distribution of the overlap will not be smooth.

```
In [3]: overlap = notebook_helpers.compute_pattern_overlap(process)
sns.distplot(overlap, kde=True, norm_hist=True, axlabel='Content overlap')
Average overlap: 0.338826769742
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x10de8ca50>
```



Code and Visualizations

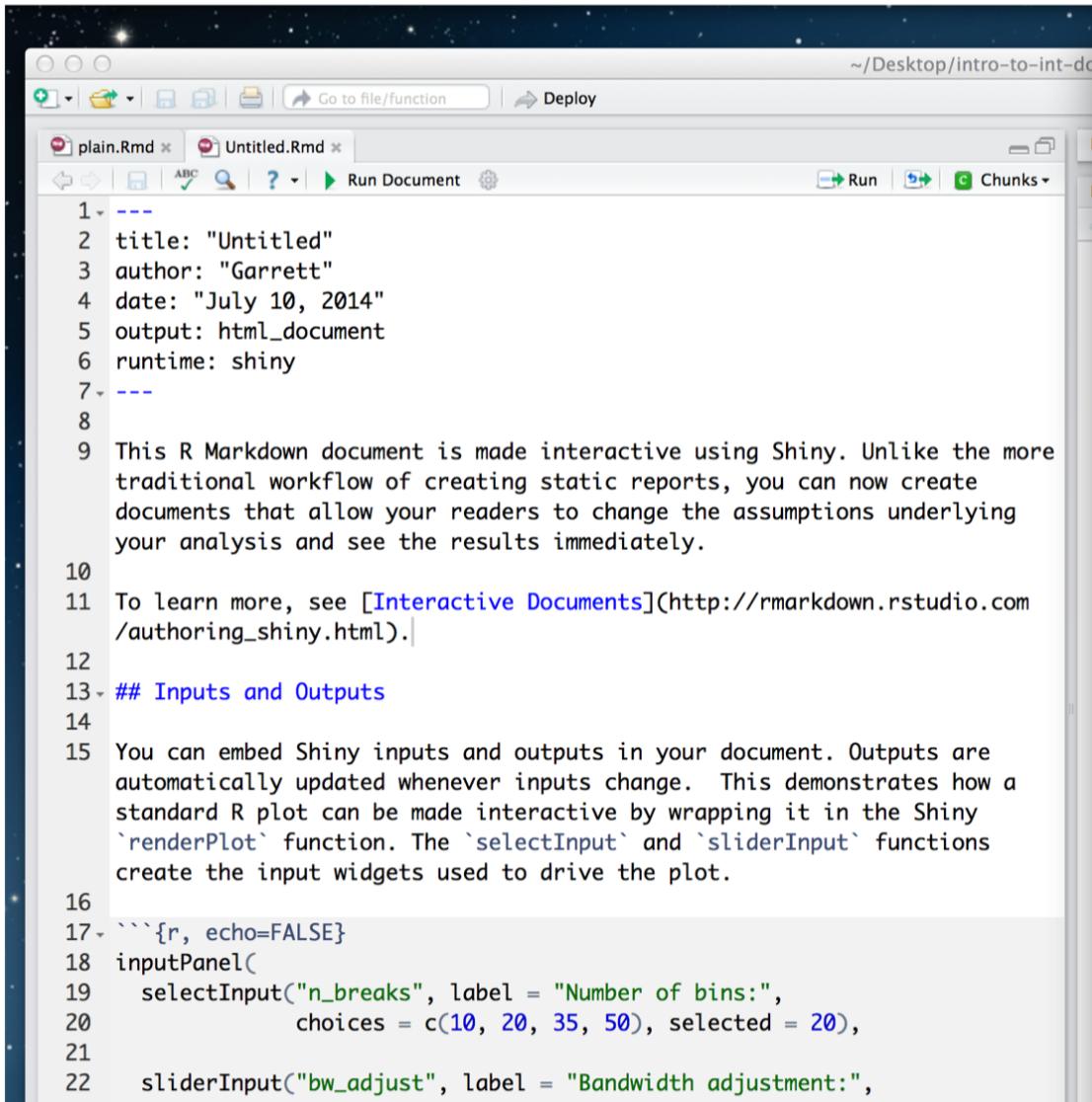
Importing external packages

Implementation of parameters

Profile plotting code

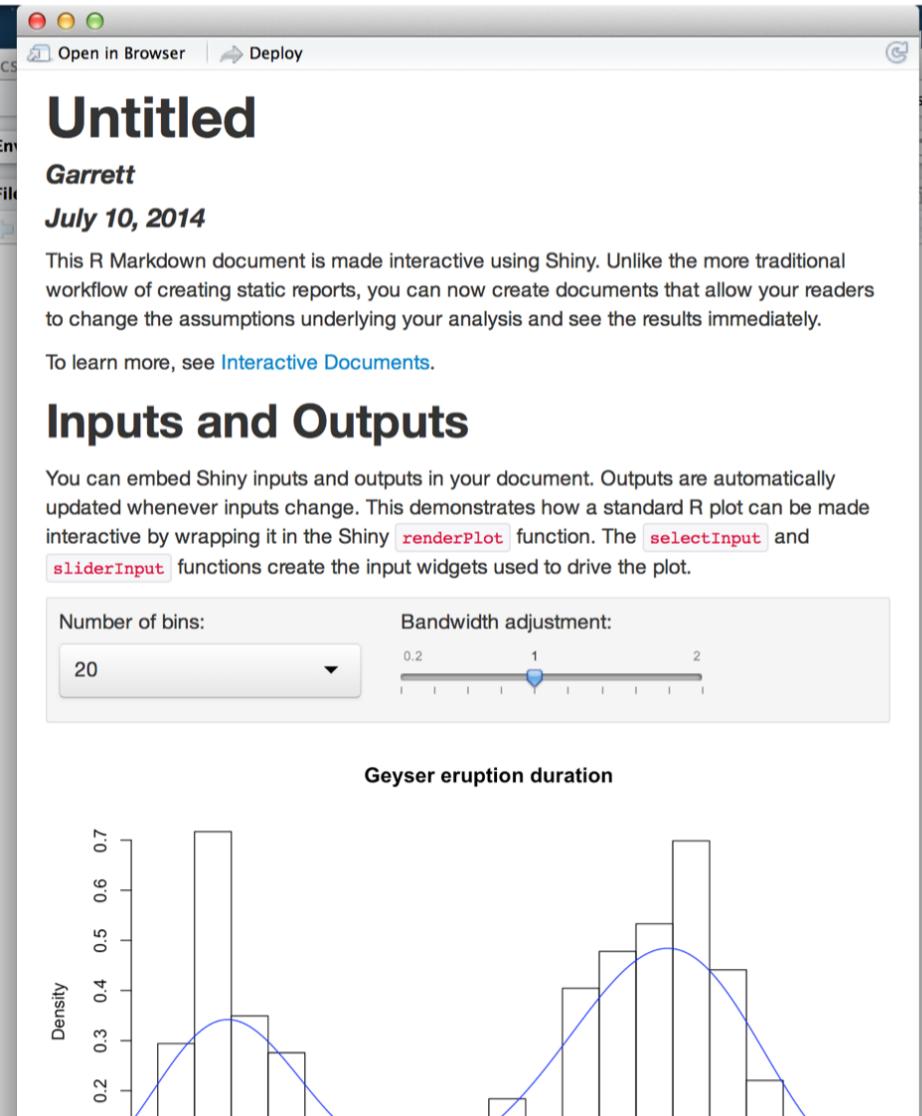
Inline plot

R Markdown Notebooks

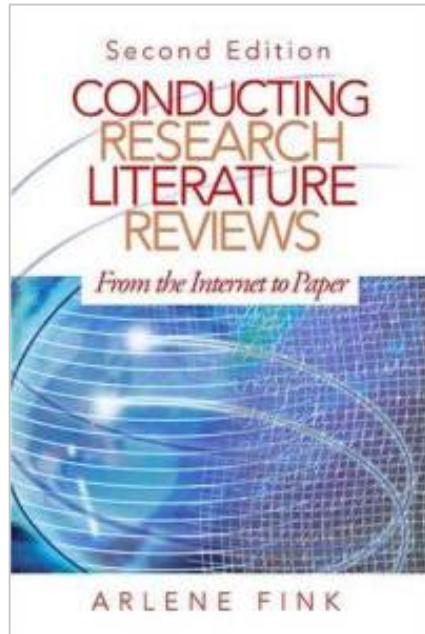


The screenshot shows the RStudio interface with two files open: 'plain.Rmd' and 'Untitled.Rmd'. The 'Untitled.Rmd' file contains the following content:

```
1 ---  
2 title: "Untitled"  
3 author: "Garrett"  
4 date: "July 10, 2014"  
5 output: html_document  
6 runtime: shiny  
7 ---  
8  
9 This R Markdown document is made interactive using Shiny. Unlike the more traditional workflow of creating static reports, you can now create documents that allow your readers to change the assumptions underlying your analysis and see the results immediately.  
10  
11 To learn more, see [Interactive Documents](http://rmarkdown.rstudio.com/authoring_shiny.html).  
12  
13 ## Inputs and Outputs  
14  
15 You can embed Shiny inputs and outputs in your document. Outputs are automatically updated whenever inputs change. This demonstrates how a standard R plot can be made interactive by wrapping it in the Shiny `renderPlot` function. The `selectInput` and `sliderInput` functions create the input widgets used to drive the plot.  
16  
17 ```{r, echo=FALSE}  
18 inputPanel(  
19   selectInput("n_breaks", label = "Number of bins:",  
20             choices = c(10, 20, 35, 50), selected = 20),  
21   sliderInput("bw_adjust", label = "Bandwidth adjustment:",
```



LITERATURE REVIEW



A literature review is a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners.

Literature reviews can serve many purposes...

- to gain an overview about a field
- to identify authors who are active in a field
- to learn about the evolution of ideas
- to overcome conflicts in the literature
- to uncover gaps in research
- to justify the relevance of research problems
- to develop an agenda / research questions for a project
- to provide background for research
- to identify the research strategies/methods that are common in an area

Types of Reviews

- **Stand-alone review** articles are research contributions that are not grounded in any original empirical data but exclusively draw from extant theoretical or empirical studies
- **Background reviews** are usually presented as parts of primary empirical articles, such as surveys and case studies; the findings from a background review are typically spread over the different sections of an article
- **Review sections (aka Related Work)** are located in between, they aim to provide a foundation for the research presented in the thesis and to motivate the research study and establish and demonstrate clearly a research gap that allows the research candidate to claim and justify the originality of their work

Example: Stand-alone Review

Alavi & Leidner/Knowledge Management



MIS Q REVIEW

REVIEW: KNOWLEDGE MANAGEMENT AND KNOWLEDGE MANAGEMENT SYSTEMS: CONCEPTUAL FOUNDATIONS AND RESEARCH ISSUES^{1, 2}

By: Maryam Alavi
John and Lucy Cook Chair of Information
Technology
Goizueta Business School
Emory University
Atlanta, GA 30322
U.S.A.
Maryam_Alavi@bus.Emory.edu

Dorothy E. Leidner
Texas Christian University
Fort Worth, Texas 76129
U.S.A., and
INSEAD
77305 Fontainebleau
FRANCE
dorothy.leidner@insead.fr

Abstract

Knowledge is a broad and abstract notion that has defined epistemological debate in western philosophy since the classical Greek era. In the past few years, however, there has been a growing interest in treating knowledge as a significant organizational resource. Consistent with the interest in organizational knowledge and knowledge management (KM), IS researchers have begun promoting a class of information systems, referred to as knowledge management systems (KMS). The objective of KMS is to support creation, transfer, and application of knowledge in organizations. Knowledge and knowledge management are complex and multi-faceted concepts. Thus, effective development and implementation of KMS requires a foundation in several rich literatures.

To be credible, KMS research and development should preserve and build upon the significant literature that exists in different but related fields. This paper provides a review and interpretation of knowledge management literatures in different fields with an eye toward identifying the important areas for research. We present a detailed process view of organizational knowledge management with a focus on the potential role of information

Example: Review / Related Work Section

Multi-Sensory Multimedia Synchronization

Niall Murray¹, Yuansong Qiao¹, Brian Lee¹, Gabriel-Miro Muntean²

¹Software Research Institute, Athlone Institute of Technology, Ireland

²School of Electronic Engineering, Dublin City University, Ireland

E-mail: ¹{nmurray, ysqiao}@research.ait.ie, ¹blee@ait.ie, ²munteang@eeng.dcu.ie

Abstract: Multi-Sensory Multimedia (mulsemedia) or Sensory experiences within multimedia are gaining interest in recent times as a potential approach to enhancing users quality of experience. However it has been shown that in different application fields that the enhancement of applications with media such as haptic, olfaction, wind, taste etc., only gains a true benefit if the congruency of the media is maintained as they exist in reality i.e. the spatially, temporally and content relations must remain intact. The key advantage of adding additional types of media is twofold: spread information processing across senses and enhancing the users perceived multimedia quality levels. In this paper, the authors introduce the reader to synchronization works involving multimedia beyond the classic audiovisual as well as reporting on the results of an experimental study which looked at user's perception of inter-stream synchronization between olfactory data (scent) and video. The results of the tests are reported and analysed considering assessor age, gender and combinations thereof.

Keywords: Multisensory Multimedia, Mulsemedia, Sensory Experience, Synchronization, olfaction

1 INTRODUCTION

Synchronization of the classic correlated media streams (audio and video) has long been an active research topic. The requirement for synchronization is to recreate temporal relationships between the various media objects at presentation time. Numerous publications report efforts to maintain the temporal relations among media streams. More recently, new demands for synchronization solutions have emerged with the advent of shared social experiences [1]. In addition, the delivery of multimedia to facilitate sensory experiences with multimedia or mulsemedia [2] brings another level of complexity to the synchronization field. Works with sensory related objects (e.g. olfaction, haptic etc.) have reported unexpected results from a synchronization perspective. For example, works reported by [3][4][6][7] indicate that assessors are quite tolerant to certain levels of inter-media skew and to lingering effects associated with olfaction. This paper is a composite of works completed to date. The interested reader can obtain detailed information on the results in

combinations to determine if differences exist between various groups. The remainder of this paper is organised as follows: Section 2 discusses related work, section 3 presents the experimental setup, section 4 presents results and the paper is concluded in section 5.

2 RELATED WORK

A fundamental requirement of any multimedia application, including those enhanced with multiple sensorial media, is synchronized display of multiple media. Research on synchronization of multi-sensory media applications is an active research area. In terms of standardization efforts, MPEG-V defines metadata representations for olfactory data among other sensory effects as part of its Sensory Effects Description Language (SEDL) within Sensory Information (part 3) [8]. Some works have evolved based on this standard, [9] that consider lighting effects, wind, heat etc. Eid et al [10] report a multi-sensory multiplexing synchronization framework. The Admux system an adaptive application layer protocol for multimedia applications incorporating haptic, visual, auditory, and scent data for non-dedicated networks. The HugMe system [11] is a synchronous haptic teleconferencing system with touch interaction to convey affection and intimacy. [12] is another work that aims to address synchronization of haptic with audiovisual media as part of an evaluation of a remote surgery application. [13] reports a synchronization mechanism to address the operational instability caused by network delay for collaborative virtual environments. [14] reports an inter-destination synchronization algorithm for haptic, audio, video streams was achieved by enhancing the virtual-time rendering (VTR) algorithm [5] to ensure fairness in a haptic building block game.

2.1 Related work: olfaction-enhanced multimedia synchronization

Little work has been documented on user perceived inter-stream synchronization of olfactory data with other media, with [6][7] for audiovisual and olfactory, [15] haptic and olfactory being the exceptions. The methodology used in these works was originally documented in [16]. In [16], inter-stream skew were artificially introduced between audio and video (lip synchronization) to determine the acceptable user perceived temporal synchronization levels. Ghosh et al. [17] published a study on the

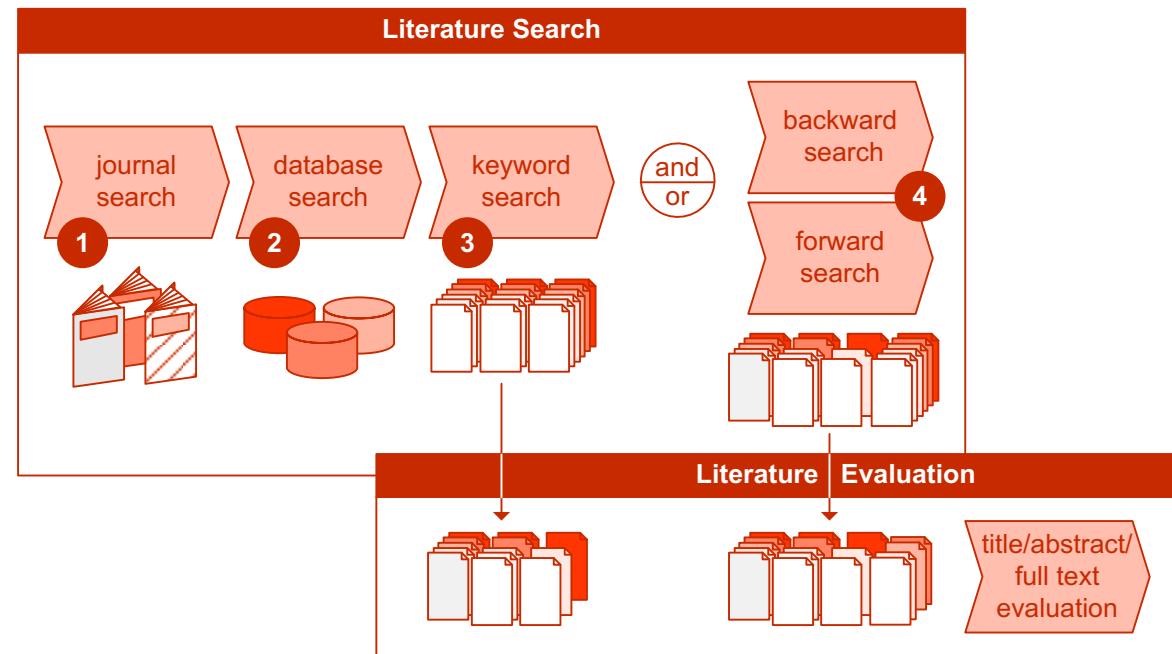
Narrative vs. Systematic Review

- From a method perspective two basic types of literature reviews are commonly distinguished: narrative reviews and systematic reviews
 - A **narrative review** (or “conventional” review) typically **draws from the reviewer’s experience in a field**.
 - It has a long tradition in research, and these reviews are written from the vantage point of the author. The narrative review thus provides a, sometimes critical, overview from the expert’s point-of-view.
 - Conversely, they have been criticized as neither exhaustive in their inclusion of source material nor objective in their viewpoints.

Narrative vs. Systematic Review

- From a method perspective two basic types of literature reviews are commonly distinguished: narrative reviews and systematic reviews
 - A **systematic** literature review, on the other hand, takes a structured approach to identifying, evaluating, and synthesizing research
 - Systematic reviews offer an **explicitly scripted and reproducible review process**
 - At the same time, systematic literature reviews have been criticized for being mechanistic and too focused on the identification of literature at the expense of scholarship in performing the actual review

Review Process



LITERATURE REVIEW

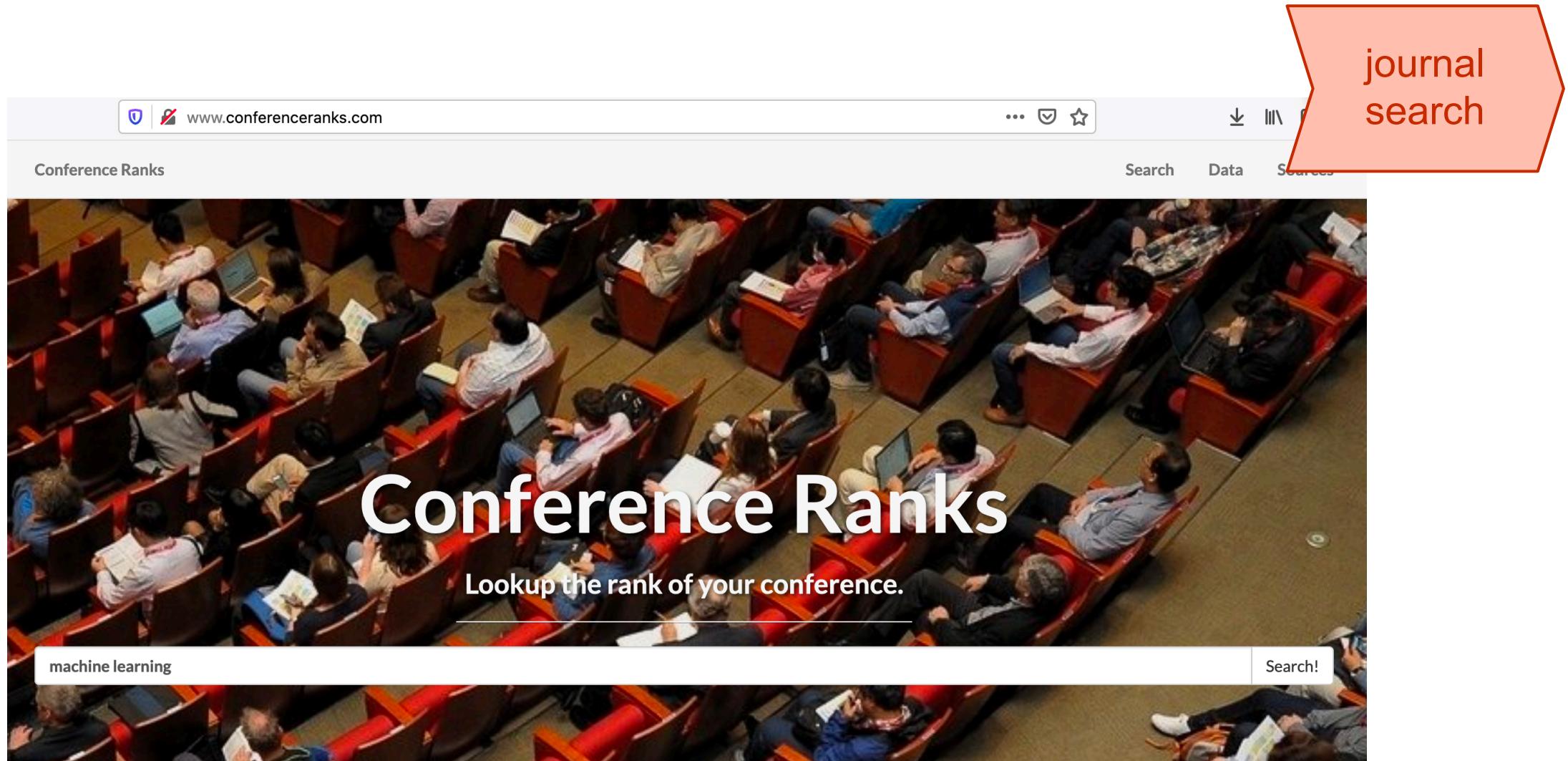
journal
search

A+ = Herausragende, weltweit führende wissenschaftliche Zeitschrift auf dem Gebiet der BWL oder ihrer Teildisziplinen										
Zeitschrift	JQ3	JQ2	Anzahl Voten	Anteil Rating o. besser	Verteilung der Voten				Nicht Wiss.	Nicht BWL
					A+	A	B	C		
Information Systems Research (ISR)	A+	A+	163	78,5%	78,5%	16,0%	3,7%	1,8%	0,0%	0,0%
Management Information Systems Quarterly (MISQ)	A+	A	201	65,2%	65,2%	27,4%	4,0%	3,0%	0,5%	0,0%
A = Führende wissenschaftliche Zeitschrift auf dem Gebiet der BWL oder ihrer Teildisziplinen										
Zeitschrift	JQ3	JQ2	Anzahl Voten	Anteil Rating o. besser	Verteilung der Voten				Nicht Wiss.	Nicht BWL
					A+	A	B	C		
Journal of Management Information Systems	A	A	111	85,6%	22,5%	63,1%	9,0%	4,5%	0,9%	0,0%
Mathematical Programming	A	A	35	80,0%	25,7%	54,3%	17,1%	2,9%	0,0%	7,9%
Journal of the Association for Information Systems (JAIS)	A	B	107	73,8%	18,7%	55,1%	17,8%	8,4%	0,0%	0,0%
Journal of Information Technology	A	B	84	67,9%	14,3%	53,6%	22,6%	9,5%	0,0%	0,0%
Proceedings of the International Conference on Information Systems (ICIS)	A	A	161	62,7%	11,2%	51,6%	24,2%	9,9%	3,1%	6,9%
Information Systems Journal (ISJ)	A	B	94	60,6%	12,8%	47,9%	30,9%	6,4%	2,1%	0,0%
The Journal of Strategic Information Systems	A	B	92	59,8%	6,5%	53,3%	32,6%	6,5%	1,1%	0,0%
European Journal of Information Systems (EJIS)	A	C	129	58,9%	9,3%	49,6%	24,0%	15,5%	1,6%	0,0%
INFORMS Journal on Computing (JOC)	A	B	54	55,6%	11,1%	44,4%	37,0%	7,4%	0,0%	0,0%
SIAM Journal on Computing	A	A	30	53,3%	16,7%	36,7%	23,3%	20,0%	3,3%	0,0%
B = Wichtige und angesehene wissenschaftliche Zeitschrift auf dem Gebiet der BWL oder ihrer Teildisziplinen										
Zeitschrift	JQ3	JQ2	Anzahl Voten	Anteil Rating o. besser	Verteilung der Voten				Nicht Wiss.	Nicht BWL
					A+	A	B	C		
Journal of the ACM (JACM)	B	B	39	92,3%	10,3%	20,5%	61,5%	7,7%	0,0%	0,0%

Title	Type	↓ SJR	H index	Total Docs. (2017)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.
1 Journal of Statistical Software	journal	13.802 Q1	103	84	262	3403	4244	262	19,99	40,51 
2 Molecular Systems Biology	journal	8.504 Q1	123	63	186	3816	1446	168	7,51	60,57 
3 Bioinformatics	journal	6.140 Q1	318	726	2360	17150	18335	2313	5,48	23,62 
4 Journal of Operations Management	journal	5.739 Q1	158	19	139	1403	843	128	5,16	73,84 
5 Wiley Interdisciplinary Reviews: Computational Molecular Science	journal	5.698 Q1	53	39	102	5480	1376	98	8,97	140,51 
6 Journal of Service Research	journal	5.247 Q1	97	29	91	1922	709	87	7,12	66,28 
7 MIS Quarterly: Management Information Systems	journal	5.085 Q1	189	59	147	4974	1214	141	7,05	84,31 
8 GigaScience	journal	5.022 Q1	24	123	150	5217	1012	122	8,27	42,41 
9 Journal of Supply Chain Management	journal	5.015 Q1	71	21	67	1676	470	63	6,10	79,81 
10 Handbook of Computational Economics	book series	4.160 Q1	20	0	13	0	20	11	0,00	0,00 
11 IEEE Transactions on Fuzzy Systems	journal	4.024 Q1	163	259	454	5986	4348	452	8,59	23,11 
12 npj Computational Materials	journal	3.826 Q1	10	54	36	3246	313	35	8,94	60,11 
13 SoftwareX	journal	3.724 Q1	5	58	52	1216	521	51	10,22	20,97 

<https://vhbonline.org/vhb4you/jourqual/vhb-jourqual-3/teirating-wi/>

<https://www.scimagojr.com/journalrank.php>



database
search

EBSCOHost

IEEE Xplore

Google Scholar

ScienceDirect

ACM Digital Library

Semantic Scholar

ProQuest

...

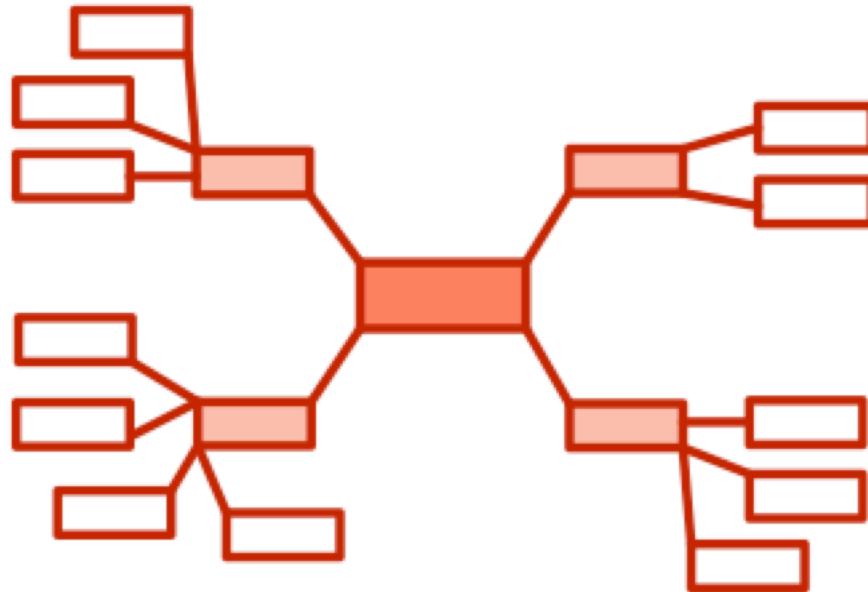
...

INFORMS Journals

AIS Electronic Library

...

keyword
search



Example

keyword
search

(("information system" OR "information systems" OR "enterprise system" OR "enterprise systems"
OR "information technology" OR "IS-enabled" OR "IT-enabled") AND ("organizational
transformation" OR "organisational transformation" OR "enterprise transformation" OR "corporate
transformation" OR "business transformation" OR "transformation of the firm" OR "strategic
transformation" OR "radical change" OR "radical transformation" OR "revolutionary change" OR
"revolutionary transformation" OR "disruptive transformation" OR "strategic change")) OR
"transformational information technology" OR "transformational information system" OR
"transformational information systems" OR "technochange" OR "transformational IT" OR
"transformational IS"

- Reference lists
- Literature/reference databases
- Google Scholar
- ...

backward
search

foreward
search

[HTML] **Beyond crowd** judgments: Data-driven estimation of market value in association football

O Müller, A Simons, M Weinmann - European Journal of Operational ..., 2017 - Elsevier

... Innovative Applications of OR. **Beyond crowd** judgments: Data-driven estimation of market value in association football ... **Crowd**-based estimation of market value ... football experts can, a style of **judgment** for which Surowiecki (2005) coined the term “wisdom of **crowds**.” Some of the ...

☆ 99 Cited by 36 Related articles All 18 versions



Evaluation

Table 2. Concept Matrix

Articles	Concepts				
	A	B	C	D	...
1		✗	✗		✗
2	✗	✗			
...			✗	✗	

Example



Table 1
Indicators of market value.

Indicator	Description	Selected references
<i>Player characteristics</i>		
Age	Age reflects players' experience and potential.	(1)–(19)
Height	Height reflects heading ability, which can influence the probability of scoring or preventing goals.	(2), (4), (11), (18)
Position	Position reflects players' flexibility on the pitch and their crowd-pulling capacity.	(1)–(19)
Footedness	Two-footedness is an advantageous footballing ability that also reflects players' flexibility.	(2), (12), (18)
Nationality	Nationality refers to a player's country or continent of birth.	(2), (6), (8), (9), (14), (16), (17)
<i>Player performance</i>		
Playing time	Playing time refers to the number of games or minutes played at the national and international levels.	(1)–(13), (15)–(19)
Goals	Goals refers to the number of goals a player has scored.	(2)–(5), (7), (8), (10)–(19)
Assists	Assists refers to the number of a player's assists that helped other players score goals.	(7), (11)–(16)
Passing	Passing refers to the number of passes to other players or the accuracy of passing.	(7), (12), (16)
Dribbling	Dribbling refers to the number and success rate of a player's ball maneuvers.	(7), (11), (16)
Dueling	Dueling refers to the number and success rate of a player's tackles, clearances, blocks, and interceptions.	(7), (12), (14), (16)
Fouls	Fouls refers to the number of fouls committed or the number of times a player has been fouled.	(7), (11), (13)
Cards	Cards refers to the number of yellow, yellow/red, and red cards received by a player.	(7), (8), (13), (18)
<i>Player popularity</i>		
News	A player's news-worthiness is reflected in press citations.	(7), (13), (14)
Internet links	Popularity is reflected in the number of links reported by web search engines like Google.	(9), (12), (13)

References: (1) Brandes and Franck (2012); (2) Bryson et al. (2012); (3) Carmichael and Thomas (1993); (4) Carmichael et al. (1999); (5) Dobson et al. (2000); (6) Feess et al. (2004); (7) Franck and Nüesch (2012); (8) Frick (2011); (9) Garcia-del-Barrio and Pujol (2007); (10) Gerrard and Dobson (2000); (11) He et al. (2015); (12) Herm et al. (2014); (13) Kiefer (2014); (14) Lehmann and Schulze (2008); (15) Lucifora and Simmons (2003); (16) Medcalfe (2008); (17) Reilly and Witt (1995); (18) Ruig and van Ophem (2014); (19) Speight and Thomas (1997)

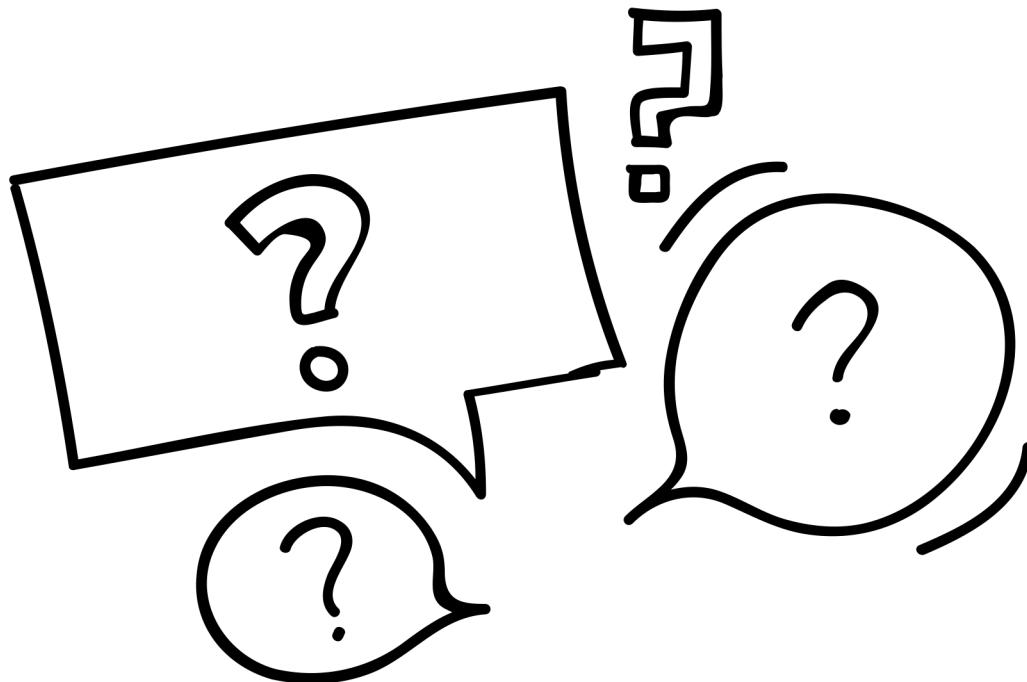
Evaluation

APA Citation Style

- In text:
 - Meier (2016) states that sentiment analysis is often used in marketing
 - Sentiment analysis is often used in marketing (Meier, 2016)
- List of references
 - Meier, R. (2016). Big Data for marketing. *Management Information Systems Quarterly*, 46 (2), 150-163.
- For more details see: <https://www.mendeley.com/guides/apa-citation-guide>

Citation Software

- EndNote (Win, Mac)
- Citavi (Win) -> free @ Paderborn University
- Mendeley (Win, Mac) -> Limit 2 GB in Cloud
- ...



Q&A

