In notitia i confide

ISKO UK Conference
25 July 2023

Martin White SearchResearch.online

Visiting Professor Information School University of Sheffield

Member of The Search Network

martin.white@intranetfocus.com

https://www.linkedin.com/in/martin-white-a7395/

https://thesearchnetwork.com/about/

The impact of information quality on enterprise search performance

ISKO UK Conference
25 July 2023

Martin White SearchResearch.online

Visiting Professor Information School University of Sheffield

Member of The Search Network

martin.white@intranetfocus.com

https://www.linkedin.com/in/martin-white-a7395/

https://thesearchnetwork.com/about/

The roadmap

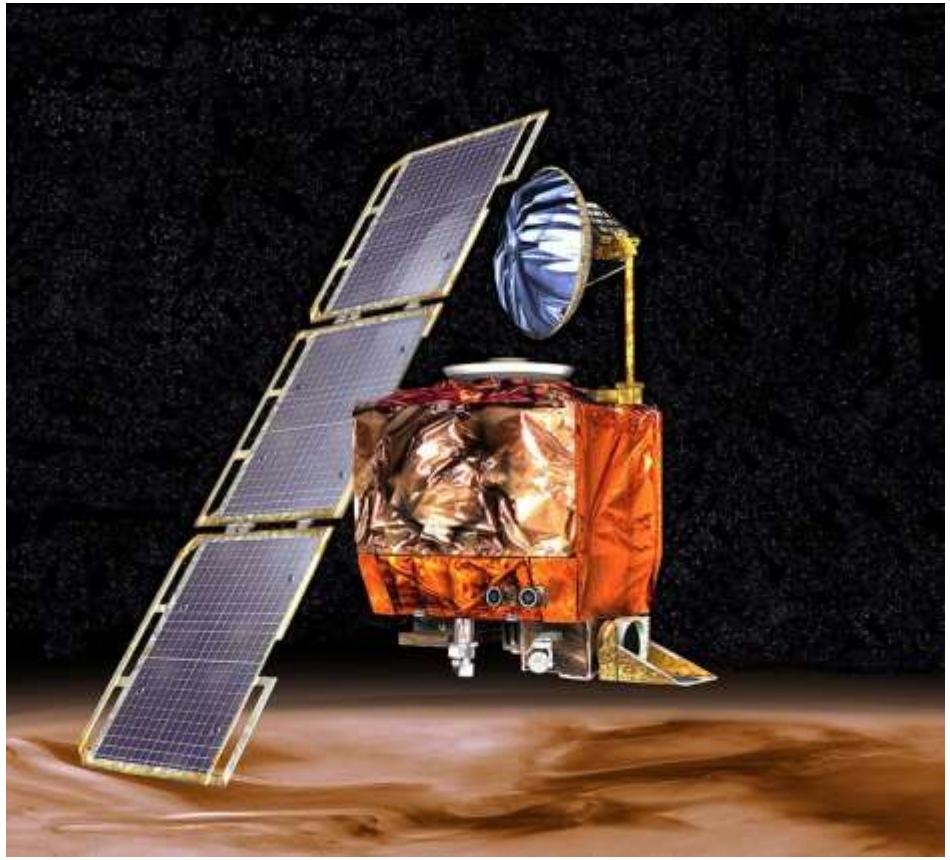

#1 Information disasters
#2 What do we mean by ‘search’
#3 Enterprise search research
#4 Inside the enterprise
#5 Search dissatisfaction
#6 Making search work

The roadmap


#1 Information disasters
#2 What do we mean by ‘search’
#3 Enterprise search research
#4 Inside the enterprise
#5 Search dissatisfaction
#6 Making search work

Imperial or metric?

Mars Climate Orbiter 1999



$500M mission cost in current value

https://en.wikipedia.org/wiki/Mars_Climate_Orbiter

https://web.archive.org/web/20010920052120/http://sunnyday.mit.edu/accidents/MCO_report.pdf

Wobbly bridge

https://www.visitlondon.com/things-to-do/sightseeing/london-attraction/bridge

Earthquake Engineering & Structural Dynamics Explore this journal >

Article

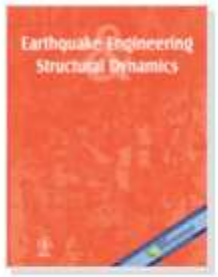
Synchronization of human walking observed during lateral vibration of a congested pedestrian bridge

Yozo Fujino, Benito M. Pacheco, Shun-Ichi Nakamura, Pennung Warnitchai

First published: September 1993 Full publication history

DOI: 10.1002/eqe.4290220902 View/save citation

Cited by: 85 articles Refresh Citing literature



View issue TOC
Volume 22, Issue 9
September 1993
Pages 741–758

Abstract

Observation of human-induced large-amplitude lateral vibration of an actual pedestrian bridge in an extremely congested condition is reported. Walking motions of pedestrians recorded by a video camera are analysed. It is found that walking among 20 per cent or more of the pedestrians on the bridge was synchronized to the girder lateral vibration. With this synchronization, the total lateral force from the pedestrians to the girder is evidently increased and it acts as a resonant force on the girder lateral vibration.

Fukushima disaster

Thatcher, A., Vasconcelos, A. C., & Ellis, D. (2015). An investigation into the impact of information behaviour on information failure: The Fukushima Daiichi nuclear power disaster. *International Journal of Information Management*, *35*(1), 57-63.

6. The Fukushima nuclear power disaster

The analysis of the reports on the Fukushima nuclear power disaster revealed three main categories of issues: cultural attitudes, information avoidance and filtering; risk management and preparation; decision-making chain and information failure. The categories highlight the interweaving between information behaviour and decisions to escalate commitment which led to the disaster. A culture had developed in the Japanese nuclear industry which was focused on maintaining an image of complete nuclear safety. Communication was informal and oral and a cost saving attitude developed in which natural disasters were viewed as low risk; therefore, resources were not provided for protective measures, causing a lack of preparedness for the disaster. Information which did not conform to pre-existing attitudes towards nuclear power was avoided, ignored or distorted. The analysis also exposes an escalation of commitment to the 'myth of nuclear safety'. To acknowledge information stating that nuclear power in Japan was unsafe would mean losing face; by avoiding and ignoring this information, the decisions made to promote nuclear power as completely safe were justified and the need to embark on costly improvement works was avoided. This was a prolonged escalation firmly establishing the avoidance behaviours which led to the disaster. These behaviours meant Fukushima Daiichi was vulnerable and when a disaster struck those involved were unable to react efficiently.

The roadmap

9

What do we mean by ‘search’?

- **WWW search** Massive volumes of everything dominated by Google and the need for $$. People want their content to be found so pay attention to quality, metadata and links
- **Web site/intranet search** Highly curated content but search is seen as an incidental add-on to the information architecture
- **Library search** Services for (primarily) academic users with highly curated content on special-purpose applications
- **E-Commerce search** Highly curated content and very good user tracking metrics where revenue improvement drives everything
- **Professional search/Systematic search** Users with excellent retrieval and subject skills working with highly curated content and a requirement for 100% recall
- **Enterprise search** Structured and unstructured content, very little of which is curated for quality because there is no incentive to make it findable

Enterprise search is different!

1. Massive amounts of structured and unstructured content (perhaps 500 million files), very little of which is curated for quality because there is no incentive to make it findable
2. Increasingly enterprise search is federated across multiple applications, and across text, data, image and video file formats
3. Many of these applications have embedded search applications
4. ‘Old’ information can be very important
5. Content in multiple languages
6. Multiple points of failure at collection, index, query and ranking level
7. We may not be searching for our personal use so past search history, recent documents and current role will be poor guides to search intent
8. ‘Relevance’ and ‘useful’ are not synonyms
9. Most searches are ‘additive’ so precision is not a metric
10. Failure to find is not an option as it puts the organisation and the employee at risk

The roadmap

12

From the archives

22 IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH—VOL. EWS-7 No. 1

Information Retrieval—
A Comprehensive Indexed
Bibliography of 1957-1961
World Literature

Machine literature search techniques have gained wide interest as solutions to the information handling problem since 1957. Feasible machine methods were used before this time, but wide acceptance and implementation has been of recent date. This subject-classified bibliography of the world's literature for the 1957-1961 period is concerned with various aspects of mechanizing storage and retrieval of documented information. Materials cited here have been drawn from sources usually exploited in large U.S. libraries, and deal with the specialized technology of documentation; including its functions of indexing, coding, staffing, systems design, and computing facilities. English, French, and German predominate as the languages of technical discourse; however the English version of a citation for a given item has been preferred to others.

S. RICHARD SIEGEL
Technical Information Center
Space & Information Systems Division
North American Aviation, Inc.
Downey, Calif.

Received May 15, 1961
[illegible] completed Dec. 15, 1961

Information storage and retrieval—commonly known as *IR*—is the function of organizing the powers of human intellect, mechanical and computing devices, and a definitive procedure into a system for filing and finding records of man's knowledge and experience. Business, government, industry, law, medicine, science and technology are joined in the common challenge of keeping up with all the information, past and present, required for effective performance in modern competitive society.

The systems approach to retrieving stored information proceeds from Dr. Vannevar Bush's statement: that, "(Man) . . . has built a civilization so complex that he needs to mechanize his records more fully if he is to push his experiment to its logical conclusions and not merely become bogged down part way there by overtaxing his limited memory. His excursion may be more enjoyable if he can reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand, with some assurances that he can find them again if they prove important." (*Atlantic Monthly*, July 1945, page 108)

This survey of the world's literature concerning the problems involved in the storing and retrieving of document information is an attempt to provide in a single corpus all materials indexed and abstracted by standard reference services during the years 1957-1961. This period was selected because it encompasses the era of the start of practical application of machine literature search systems to the technical information retrieval problem. Elements treated during this period include indexing, classification and coding schemes; systems proposals; plans for cooperation in using information; planning of information facilities; programs for training personnel to carry out the manifold aspects of operating IR systems; and, an impressive amount of criticism and evaluation.

Most of the literature surveyed treats the specialized technology of documentation which has been variously described as the science of ordered presentation and preservation of records of knowledge involving collecting, storing, and organizing bibliographic information for optimum access and specificity; or as the art of facilitating the use of recorded specialized knowledge through its presentation, reproduction, publication, dissemination, acquisition, and retrieval. The literature also shows that the most sophisticated systems in operation are used in the fields of chemistry and metallurgy.

Universal interest in this world-wide problem is evidenced by the variety of sources and the wide range of languages represented in this indexed bibliography.

ELECTRONIC DIGITAL MACHINES
FOR HIGH-SPEED INFORMATION SEARCHING

by

Philip Rutherford Bagley

1951

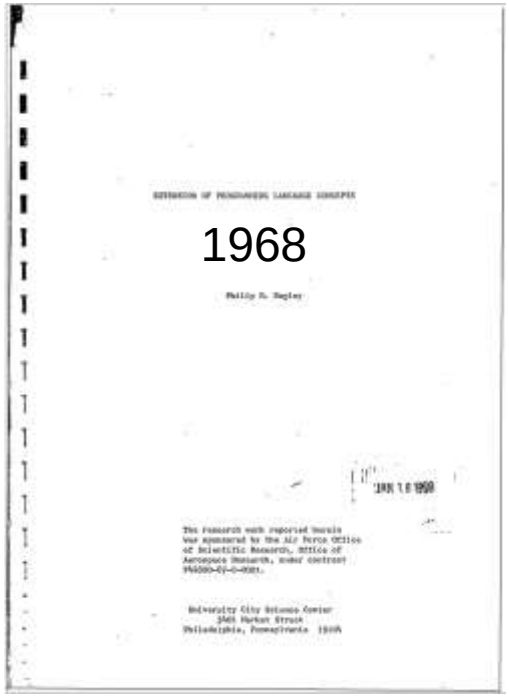
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
(1951)

Signature of Author ...
Department of Electrical Engineering, August 24, 1951

Certified by .....
.....
Thesis Supervisors

.....
Chairman, Departmental Committee on Graduate Students

Electronic digital machines
For high-speed information
searching



EXTENSION OF PROGRAMMING LANGUAGE CONCEPTS

1968

Philip R. Bagley

JAN 10 1969

This research was reported herein was sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, under contract F44620-67-C-0021.

University City Science Center
3401 Market Street
Philadelphia, Pennsylvania 19104

Extension of programming language
concepts

https://stevenbagley.net/blog/philip-r-bagley.html

The birth of enterprise search

The LEADER retrieval system

by DONALD J. HILLMAN and ANDREW J. KASARDA

Lehigh University
Bethlehem, Pennsylvania

INTRODUCTION

The LEADER system is a new service-oriented prototype designed to meet the retrieval needs of research scientists working within or in conjunction with the Center for the Information Sciences at Lehigh University. In the first part of this paper, we describe the major conceptual apparatus and principal design features of LEADER, while the second part contains a brief discussion of system implementation and user interaction.

The name "LEADER" is an acronym for "**LE**high **A**utomatic **D**evice for **E**fficient **R**etrieval," and is thus similar to other acronyms in possessing both an intended meaning as well as an actual referent. Imaginative readers can undoubtedly supply alternative and presumably more ribald interpretations of the same six characters, but this is rather incidental to the main goal of the LEADER system, which is to provide a very highly user-oriented facility for the negotiation of open-ended inquiries and interactive browsing. To help meet this objective, the system includes on-line processing of requests, using a novel and relatively inexpensive hardware configuration, and serially organizes its output in the form of document references, citations to documents, and complete textual passages selected from one or several documents, in any way that the user specifies. This ability of the user to control output is but one feature of an overall interactive procedure which begins when an initial request is entered into the LEADER system in the form of a set of sentences describing the user's problem. Each input sentence must, of course, be grammatically well-formed, but there is no restriction on vocabulary. A typical inquiry might read:

> "I would like to know whether modular bounded functionals have ever been used in theoretical studies of retrievable sets, and if so by whom and with what results. If there has been no application of this type, I would be interested to learn of any work in retrieval theory that makes use of Borel functions. If there is no such work, please direct me to retrieval studies involving topological measures or metric spaces in general."

Inquiries such as these are presented directly to the system and displayed on a CRT scope. As each inquiry is displayed, it is also automatically analyzed by the same procedures used to process the full text of input documents. That is to say, LEADER treats both documents and queries as entities of the same logical type to begin with, so that the logical and referential structure of an inquiry is accorded just as much importance as the structure of a document. The goal of text processing is therefore the same throughout, *viz.*, to determine what each group of input sentences is *about*, whether they constitute a document or an inquiry, and to establish major patterns of conceptual relatedness between documents and terms used either in document or query characterization. The text-processing features of LEADER thus include elements of syntax, semantics, and logic.

After the sentences of an initial inquiry have been analyzed into concept-denoting expressions and their logical interrelationships, LEADER is able to fashion an appropriate response to the user's retrieval needs by comparing the conceptual structure of the inquiry with the general structure of the data base. This comparison is conducted *via* a man/machine dialogue in which LEADER instructs and interrogates the user, attempting to acquaint him with the nature of its stored information so that each inquiry can be negotiated through successive modifications of the user's stated interests. The dialogue itself is carried out on a CRT scope.

The user may call for document references, citations, or passages of relevant text at any time during the negotiation, so that by a process of selective browsing

447

Interactive retrieval

LEADER is designed to encourage user interaction with the structured material of a corpus of scientific or managerial data so as to maximize the influence of information flow on decision making. The data entry procedures are sufficiently general to accommodate several different types of data base, provided only that each consist of well-formed English sentences. In addition, the response capability is flexible enough to permit retrieval ranging from the enumeration of simple bibliographic data, on the one hand, to full-text display, on the other.

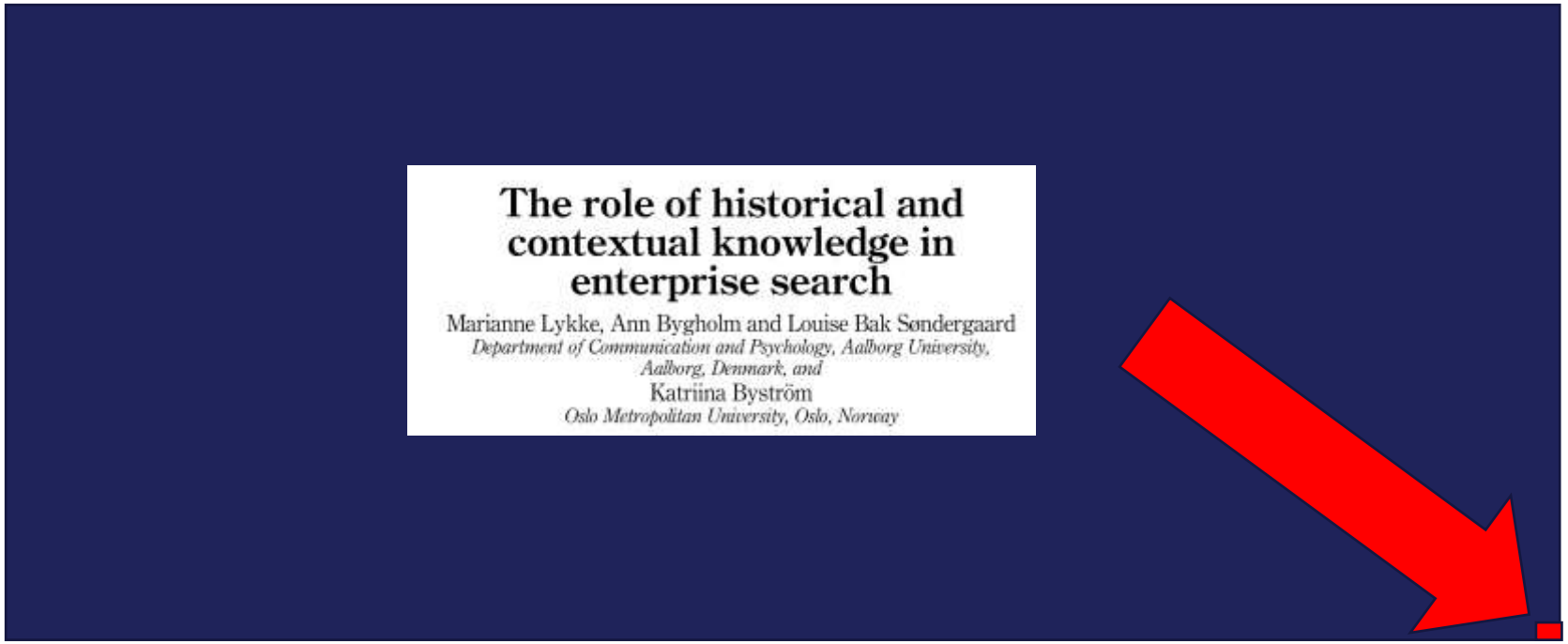
The extent to which information flow contributes to decision-making is certainly affected by the ability of an information system to adapt itself to a user's needs.

APA Hillman, D. J., & Kasarda, A. J. (1969, May). The LEADER retrieval system. In *Proceedings of the May 14-16, 1969, spring joint computer conference* (pp. 447-455).

The state of enterprise search research

IR Anthology

The IR Anthology currently contains 62846 papers on the study of information retrieval.



Presentation content validation

- Outcomes of around 40 enterprise search projects and a similar number of intranet projects, all of which had a search element
- Participation in search-related conferences in the UK, USA, Denmark, Sweden and Germany
- Authorship of three of the four books on enterprise search

- Experience gained from The Search Network members
- Questions asked at my lectures at the Information School, University of Sheffield and at City University since 2002

The roadmap

17

Tasks….

Saastamoinen, M., & Järvelin, K. (2017). Search task features in work tasks of varying types and complexity. *Journal of the Association for Information Science and Technology*, *68*(5), 1111-1123.

….or decisions?

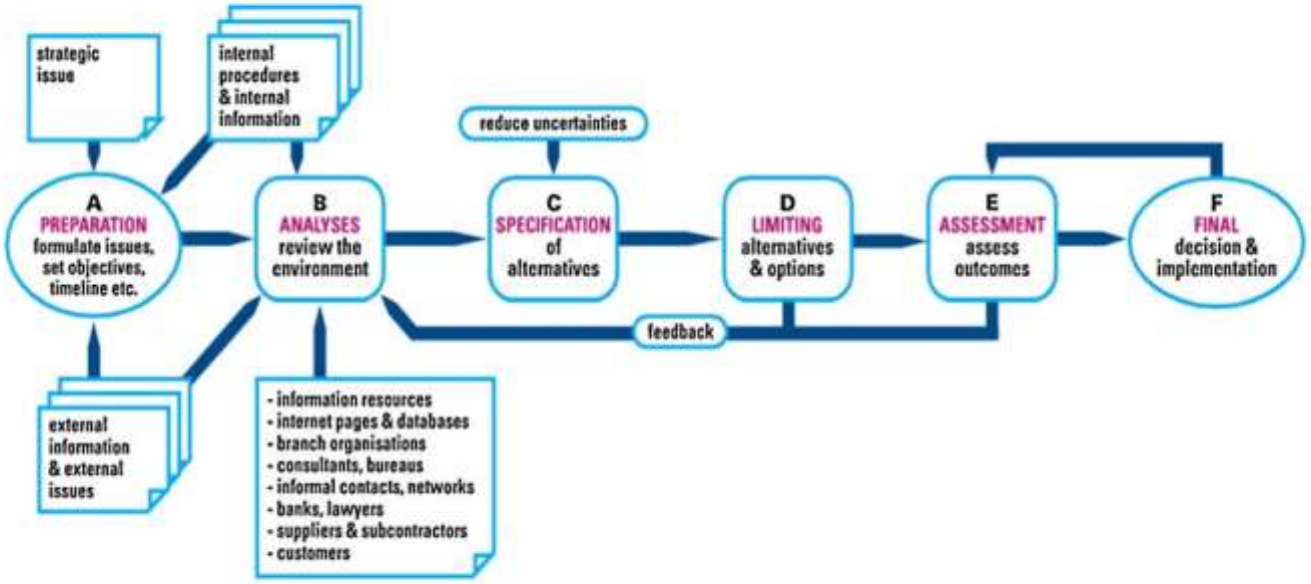
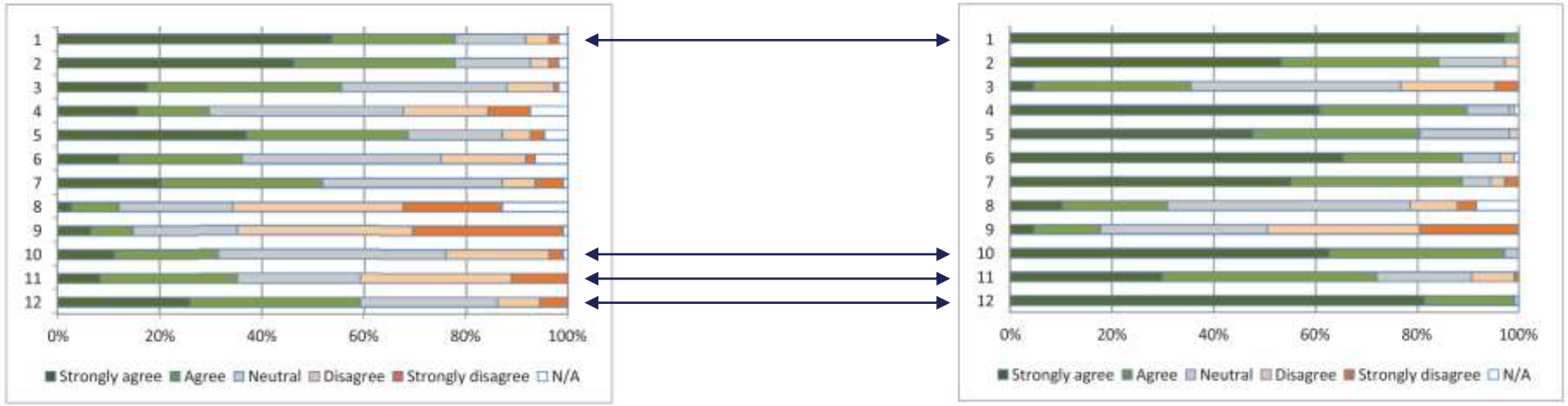


Fig. 1. Model of the phases of a rational decision-making process. Rounded boxes indicate the five phases in the decision process; square boxes contain parameters that provide input for the indicated actions. Arrows indicate the main direction of interactions.

End-user enterprise personas

A1: Legal

A3: Healthcare



Likert diagrams for legal, recruitment, healthcare and patent sectors showing the value placed on search functionality during query formulation, including Boolean logic (1), proximity operators (2), relevance ranking (3), truncation (4), wildcards (5), field operators (6), query expansion (7), query translation (8), case sensitivity (9), abbreviations (10), misspellings (11) and synonyms (12). The patent survey did not include 9–12.

Information retrieval in the workplace: a comparison of professional search practices — University of Strathclyde

Information quality dimensions

Table 3 Representational information quality and its characteristics (*N* = 34)

Characteristic	*Exemplary quotation*
Conciseness 26 (76.5%)	*The way information is presented is not succinct. I always believe it could be expressed with the half of the words they use.*
Presentation 24 (70.6%)	*Information is presented in an inexplicable way.*
Understandability 14 (41.2%)	*The information I need for my work is formatted such that it requires extra effort to understand it.*

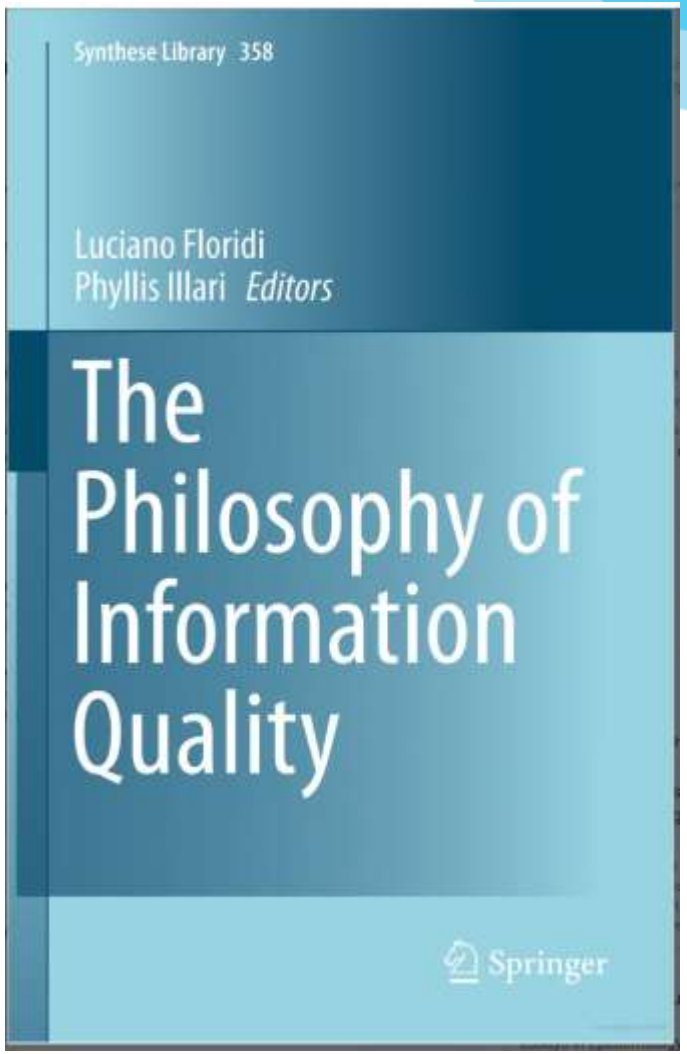
Note: Total number of interviews the characteristic was mentioned in percentage in relation to *N* = 34).

Table 4 Contextual information quality and its characteristics (*N* = 34)

Characteristic	*Exemplified quotation*
Completeness 15 (44.1%)	*The information provided is not complete. I lack explicit information that helps me to prepare my sales talks.*
Relevance 27 (79.4%)	*The information provided is not relevant to my task. It does not make it easier for me to decide which steps to follow and which department I should forward the customer request to.*
Timeliness 12 (35.3%)	*I believe the information is not provided in time. We need to answer customer questions about new products but we are not provided with the information when the new product is launched.*
Usefulness 28 (82.4%)	*The information is simply not usable for our task as there is so much information presented that we cannot use it.*

Note: Total number of interviews the characteristic was mentioned in percentage in relation to *N* = 34).

Laumer, Sven, Christian Maier, and Tim Weitzel. "Information quality, user satisfaction, and the manifestation of workarounds: a qualitative and quantitative study of enterprise content management system users." *European Journal of Information Systems* 26.4 (2017): 333-360.



Governance, risk and compliance

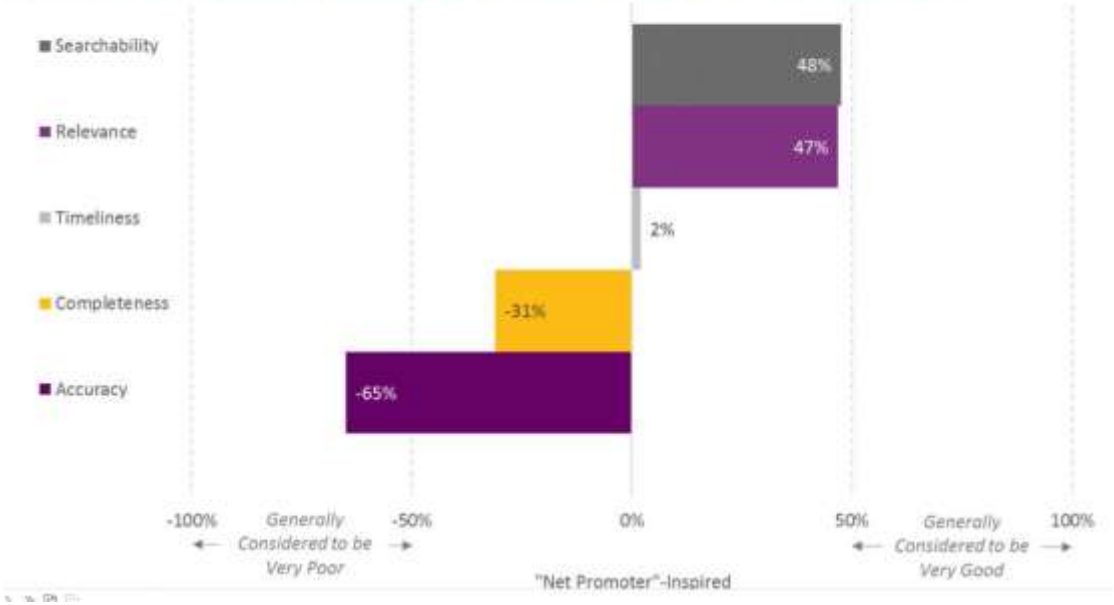
ID	Quality Factor	Example of Quality Metric
Intrinsic GRC DQ		
S1	Incomplete data; missing attributes	Percentage of missing attributes per control
S2	Heterogeneous levels of granularity	Difference between attribute length and average attribute length
S3	Duplicate content	Number of attributes with identical content
S4	Contradicting content	Number of attributes with similar but contradicting content
S5	Spelling and grammar	Number of spelling/grammar mistakes
S6	Abnormally short attributes	Length of description below threshold
S7	Comprehensibility	Number of incomprehensible requirements

Christian Sillaber, Andrea Mussmann, and Ruth Breu. 2019. Experience: Data and Information Quality Challenges in Governance, Risk, and Compliance Management. *J. Data and Information Quality* 11, 2, Article 6 (March 2019), 14 pages.
https://doi.org/10.1145/3297721

Good review of the literature!

Quality of information assessment

Figure 3: Accuracy, Completeness, and Timeliness are Currently the Most Pressing Issues with Your Quality-of-Information

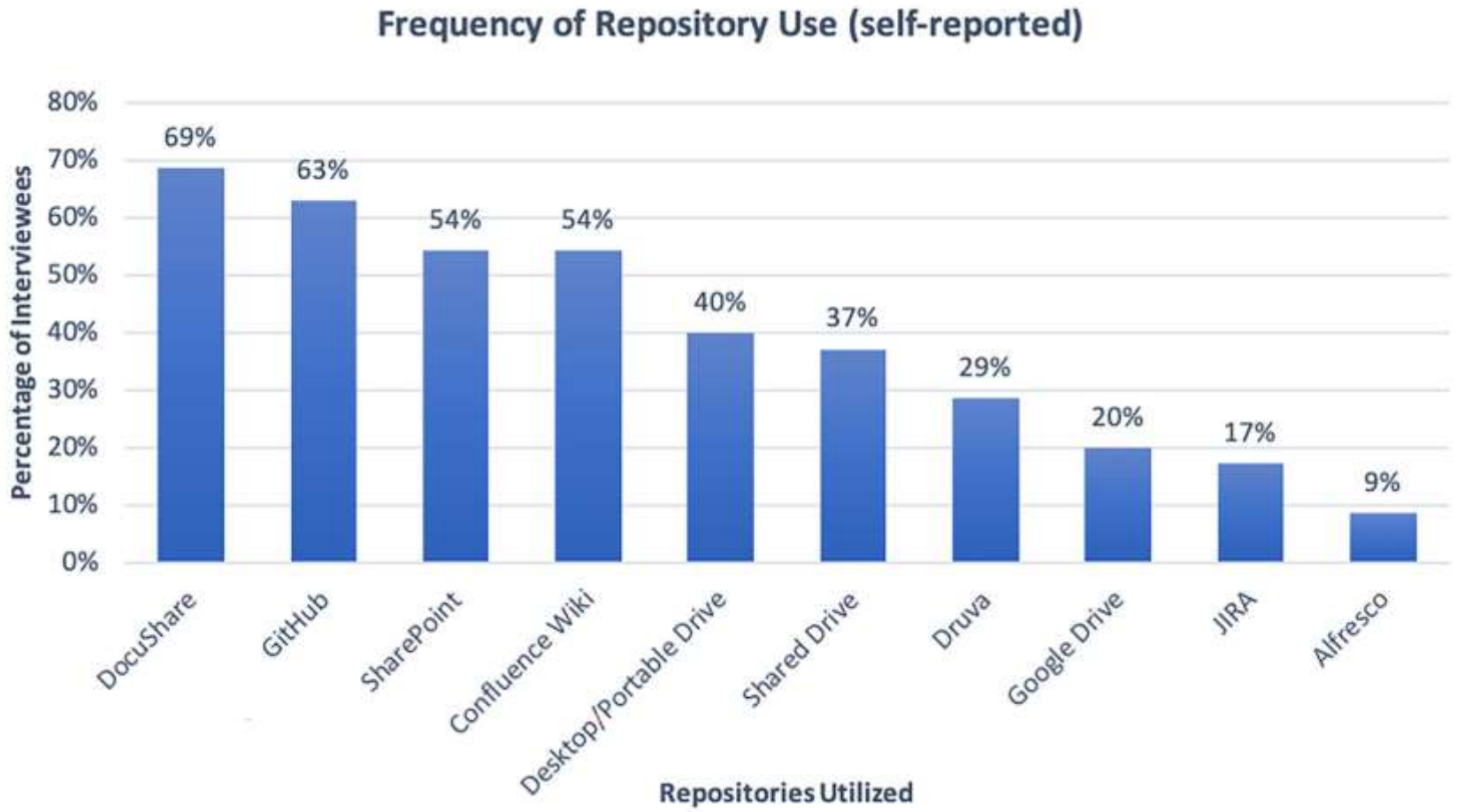


Source: Aberdeen Group, September 2017

Your enterprise decision-makers have access to relevant, timely, complete, and accurate information **between just 7% and 43% of the time**, with a **median likelihood of only 23%**. *This* is why decision-makers continue to rely so heavily on mere intuition and gut feel!

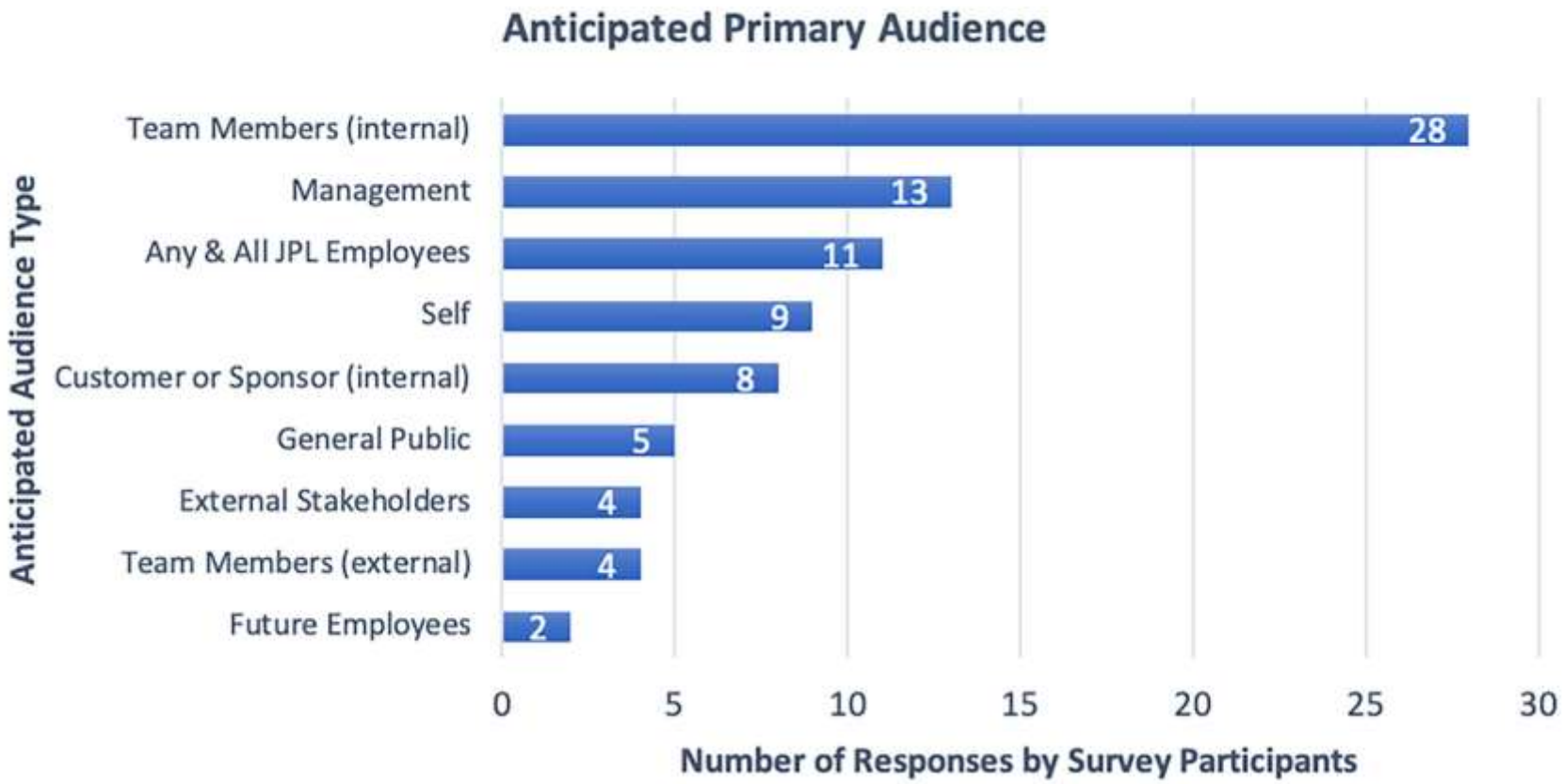
Do you really want your important business decisions to be made based primarily on personal intuition and judgment calls, more than 75% of the time?

Jet Propulsion Laboratory case study



Mathieu, Camille. "Defining knowledge workers' creation, description, and storage practices as impact on enterprise content management strategy." *Journal of the Association for Information Science and Technology* 73.3 (2022): 472-484.

Writing for friends (JPL case study)



What employees are looking for

European biotech company with 7500 employees
Just under 300,000 queries in a four-month period
Approximately 3750 queries a day of which 2400 were people searches

Top category (%)	Highly frequent (n = 479)
People search	63.4
Quality	4.4
IT	2.6
HR	2.2
Product	1.9
Finance	1.1
Facility services	1.0
Sourcing	1.3
Intranet	0.8
No top category	17.2

The role of historical and contextual knowledge in enterprise search

Marianne Lykke, Ann Bygholm and Louise Bak Søndergaard
Department of Communication and Psychology, Aalborg University, Aalborg, Denmark, and
Katriina Byström
Oslo Metropolitan University, Oslo, Norway

Content granularity

Annex

Table 1: IDB Knowledge products included in the Brik database

Type	Description
Annual Reports	Include reports published yearly to provide information about the Bank's work and/or some development issue during a year. IDBG annual reports include the Bank's official Annual Report, Development in the Americas (DIA), the Development Effectiveness Overview, as well as yearly reports related to OII, ICIM and other Bank Departments and funds.
Books	Include books published under the responsibility of EXR that originate in the Board of Governors, Board of Executive Directors (including books written by members of the Board of Governors or Board of Executive Directors in their capacity as such), the Office of the Presidency, and the Office of the Executive Vice President or published by other departments, including EXR, following the quality control and publication procedures described in the "AM-331 Procedures for the Publication of Knowledge Products".
Catalogs and brochures	Include publications designed to provide the public with information of Bank programs, results, goals, or identity and on relevant data or issues in the region (EXR is responsible for the publication of any corporate-level catalog or brochure. For any other content, the corresponding Manager authorizes the publication of Catalogs and Brochures).
Co-publications	Include publications developed and/or published by the Bank along with external organizations (such as other international organizations or academic institutions).
Databases and datasets	Include applications that combine back-end data with a front-end web-based interface where users can query and manipulate data; and, Datasets, which are raw data files, usually accompanying papers. Data contents of any Databases & Datasets shall be approved by the Department Manager. User interface for display of Databases shall be approved by EXR Departments interested in publication of Databases shall contact EXR.
Discussion Papers	Include documents shared with a community of specialists and policy makers (both within the Bank and with external community members). Discussion Papers shall be approved by a Division, Unit Chief, or Country Department Managers in case of VPC.
Learning materials	Include supporting documents for learning events, such as courses or workshops (study guides, class presentation and lectures, class reading materials, audiovisual media, simulations, case studies, methodological guidelines, teaching cases, etc.). Learning materials should generally serve a well-defined pedagogical objective; notes or stand-alone slide presentations (such as PowerPoint) from a learning event or conference would in most cases not qualify as meeting such an objective.
Magazines, Journals and Newsletters	Include periodic publications aimed at disseminating specialized information to a general audience (for example IDEA and the INT/INTAL newsletter). The corresponding Department Manager shall authorize the publication of the Magazine, Journal, or Newsletter.
Monographs	Include briefs on a single subject prepared for targeted audiences or specific purposes, such as presentations at events and policy meetings. They are papers of usually less than 100 pages that are often designed, typeset, and printed in limited quantities, depending on their use.
Policy Briefs	Include publications that discuss a development policy issue and outline courses of action, including specific policy recommendations. The author of a Policy Brief shall receive the technical approval of his or her corresponding Manager. All Policy Briefs shall be produced in a single series whose numbers shall be assigned via a web-based system, available through the KP Procedures Intranet Site. The research from which the Policy Brief is developed shall be produced under the auspices and/or leadership of the Bank (that is, the Policy Brief series shall not be a channel for disseminating the academic production of other institutions).
Technical Notes	Include a wide range of sector notes, good practices, project and other evaluations/reviews, documentation of lessons learned, case studies, methodological notes, and other documents of a technical nature. Technical Notes¿ primary audiences are government officials and other development practitioners. Any Technical Note may be published for external distribution with approval of the Division Chief, Unit Chief, or Country Department Manager in case of VPC. These KPs shall be produced in a single series whose numbers shall be assigned via a web-based system.
Working Papers	Include documents prepared for disseminating research and survey studies that, while conforming to rigorous research standards, need not be a final product, as their purpose is not only to inform but also to stimulate discussion. The audience for Working Papers is largely academic but may also include policymakers and private sector professionals. In keeping with the standards applied by most academic journals, Bank Working Papers shall not exceed 60 pages.

Source: OVE based on BRIK database

https://publications.iadb.org/en/approach-paper-knowledge-generation-and-dissemination-inter-american-development-bank-group

The Babel dimension

Language	Content items as % of total	% speaking the language as their primary language
English	73	24
German	13	25
Spanish	4	11
Portuguese	3	4
Japanese	2	6
Italian	2	5
French	1	6
Chinese	1	4
Polish	1	3

Source – client of Intranet Focus Ltd

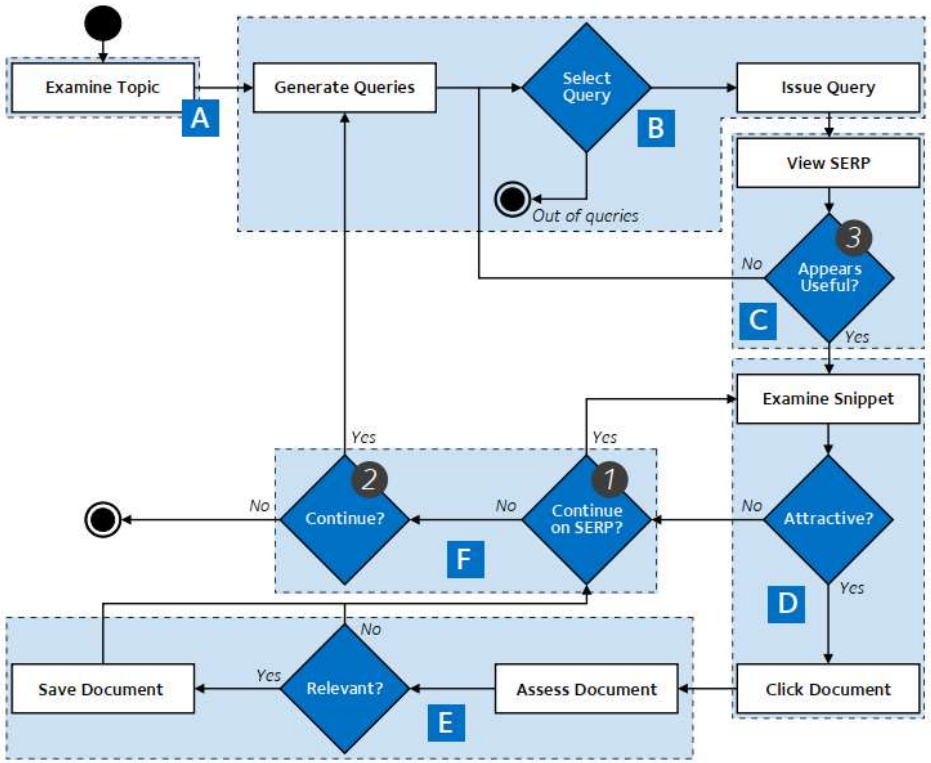
76% of employees are searching English-language documents in their second, or maybe third language

Very little work has been carried out on either the impact of NLP on search performance or on how best to present results in multiple languages

Harvey, Morgan, and David Brazier. "E-government information search by English-as-a Second Language speakers: The effects of language proficiency and document reading level." *Information Processing & Management* 59.4 (2022): 102985.

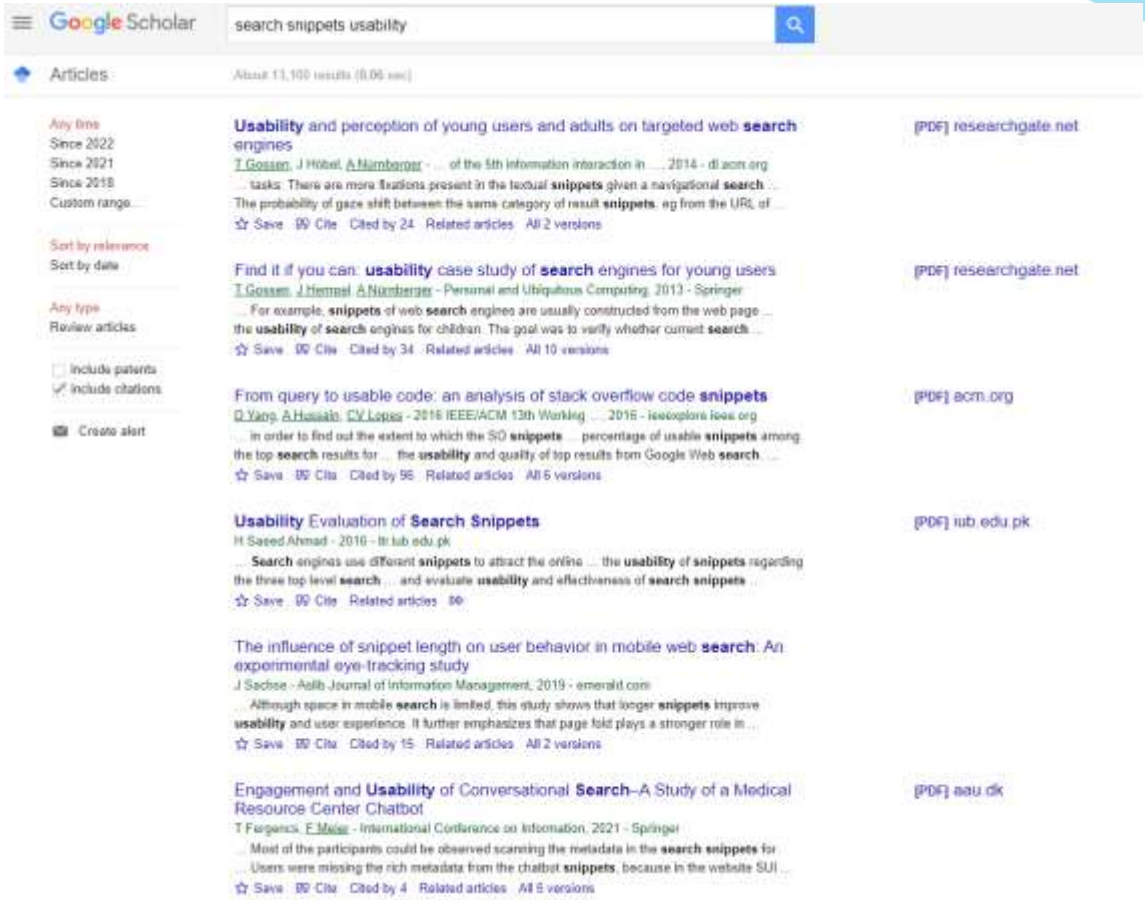
Local vs ‘corporate’ language

Snippets are of crucial importance!



Maxwell, David Martin (2019) *Modelling search and stopping in interactive information retrieval*. PhD thesis.

https://theses.gla.ac.uk/41132/



The impact of mis-spellings

MEDICAL REFERENCE SERVICES QUARTERLY
2023, VOL. 42, NO. 3, 211–227
https://doi.org/10.1080/02763869.2023.2214038

Common Misspellings and Their Impact on Health Sciences Literature Search Results

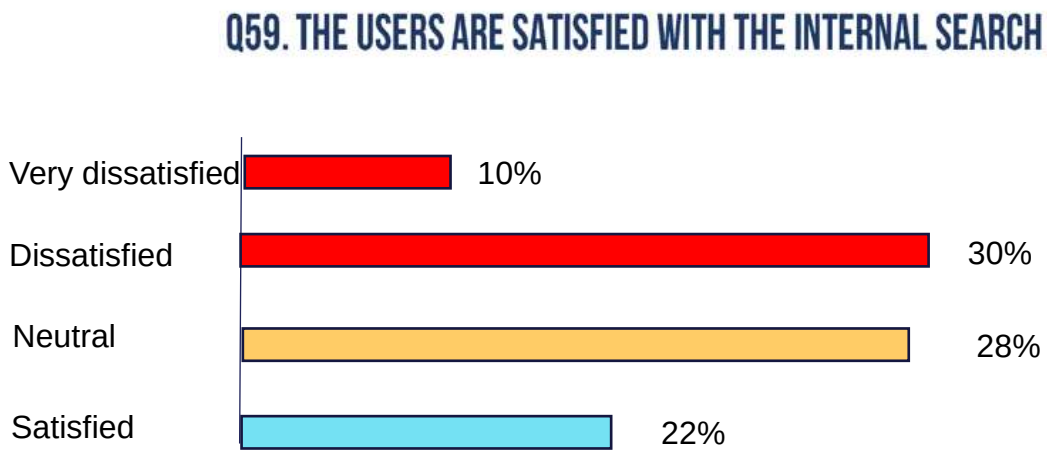
The inability to retrieve articles because of misspelled words has obvious implications for patient care if information vital to treatment cannot be found. Furthermore, missed citations have an implication for evidence-based practice (e.g., systematic reviews that aim to answer a clinical question based on evidence synthesized from all relevant research studies) as there would be an added burden on searchers to include misspellings in their search strategies.

The roadmap


#1 Information disasters
#2 What do we mean by ‘search’
#3 Enterprise search research
#4 Inside the enterprise
#5 Search dissatisfaction
#6 Making search work

"Okay, Houston, we've had a problem here”

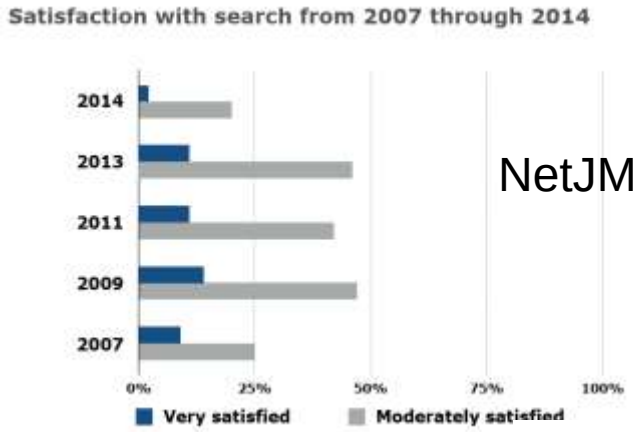
Apollo 13 13 April 1970 055.35.55



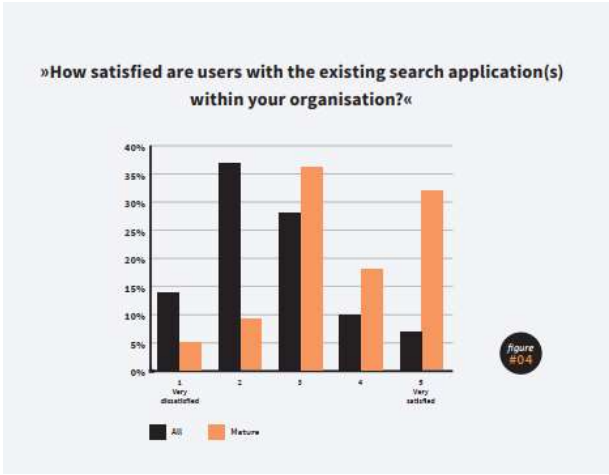
Based on benchmark reports from 230 Scandinavia organisations

IntraTeam Search Benchmark Survey report 2020 © www.intrateam.com

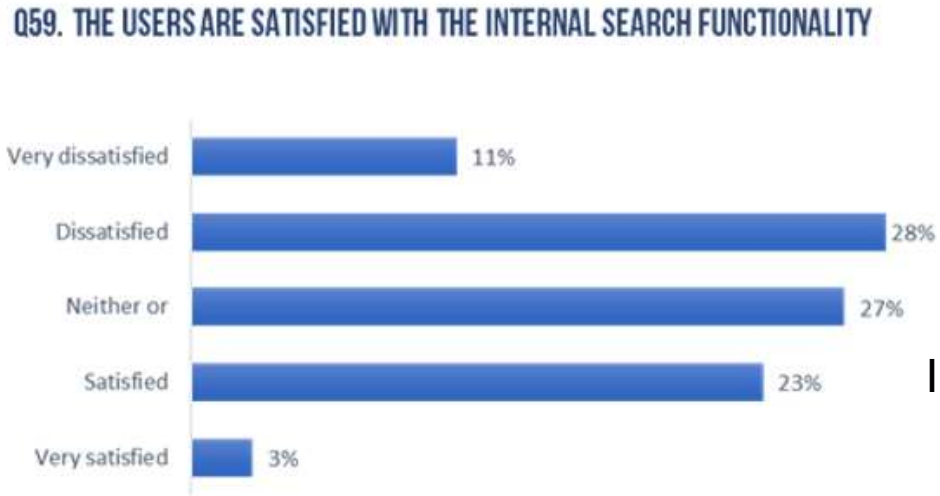
Search dissatisfaction 2007 - 2021



NetJMC 2007-2014



Findwise 2017



IntraTeam 2021

The roadmap


#1 Information disasters
#2 What do we mean by ‘search’
#3 Enterprise search research
#4 Inside the enterprise
#5 Search dissatisfaction
#6 Making search work

The challenge

Correspondence

Filippo Menczer, David Crandall, Yong-Yeol Ahn & Apu Kapadia
Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA.

https://doi.org/10.1038/s42256-023-00690-w

Addressing the harms of AI-generated inauthentic content

Generative AI tools lower the cost of generating false but credible content at scale[1], defeating the already weak moderation defenses of social media platforms. Using inauthentic accounts and other tricks to exploit algorithmic and socio-cognitive vulnerabilities, bad actors can monetize false and harmful content, commit fraud and manipulate opinions for political gains[2,3]. Generative AI tools such as ChatGPT make it easier to create large volumes of false (but convincing) social media profiles and content. Narratives can even be tailored to a particular community by an inauthentic influence campaign. For example, through health disinformation, a foreign adversary can make an entire population more vulnerable to a future pandemic[4].

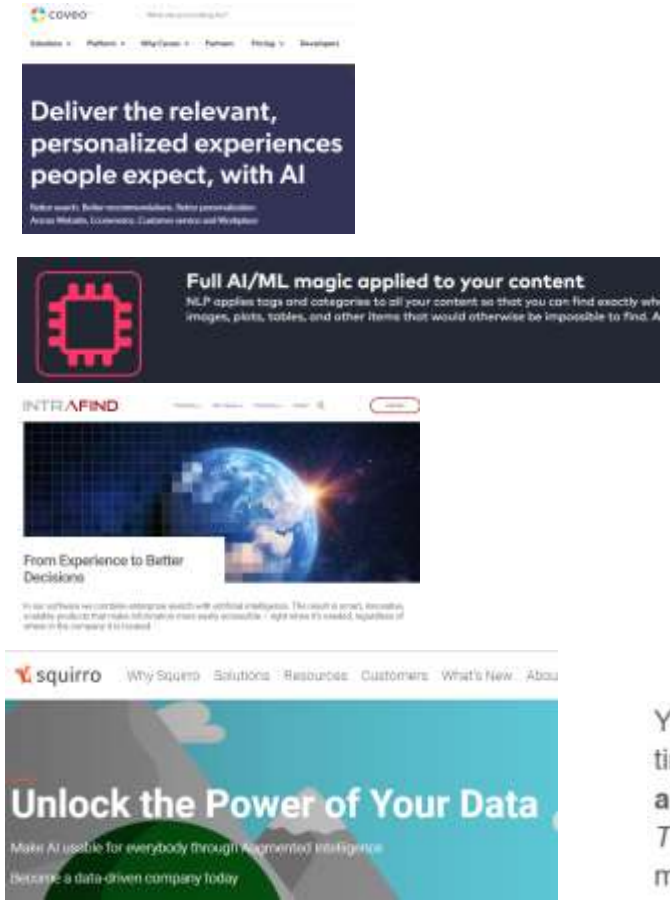
The search for solutions to mitigate these adversarial threats creates significant research challenges. A silver-bullet approach is unlikely to exist. Therefore, multiple research directions in machine intelligence should be explored in parallel, focused on detection, moderation and regulation.

An intelligence perspective

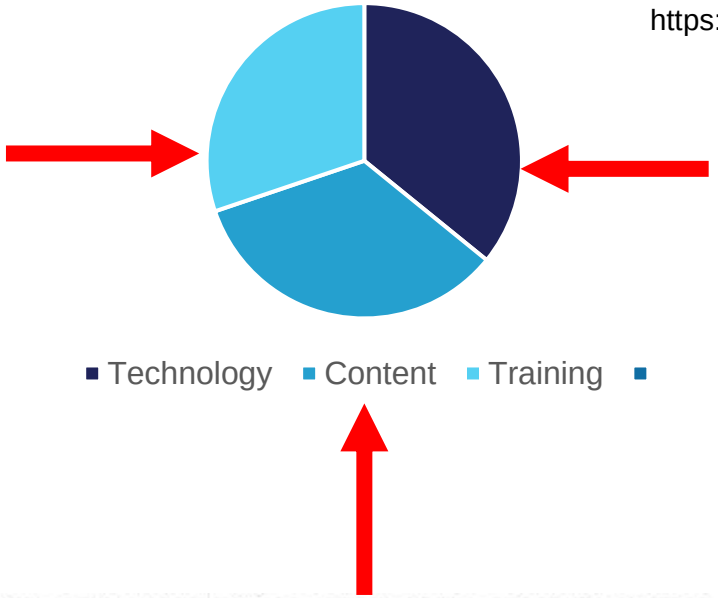
Current LLMs show promising potential as basic productivity assistants to improve efficiency of some repetitive intelligence tasks. But the most promising use cases are still on the horizon, and future efforts should focus on developing models that understand the context of the information they are processing – rather than just predicting what the next word is likely to be.

https://cetas.turing.ac.uk/publications/large-language-models-and-intelligence-analysis

Making search work



Causes of search dissatisfaction



Your enterprise decision-makers have access to relevant, timely, complete, and accurate information **between just 7% and 43% of the time**, with a **median likelihood of only 23%**. *This* is why decision-makers continue to rely so heavily on mere intuition and gut feel!

Source: Aberdeen Group, September 2017

Cleverley, P. H., & Burnett, S. (2019). Enterprise search and discovery capability: The factors and generative mechanisms for user satisfaction. Journal of Information Science, 45(1), 29–52. https://doi.org/10.1177/0165551518770969

The PIBP implies that users' information behavior goes much deeper than understanding how to use the software. It implies that the users' knowledge of how the software actually works (in this case: their knowledge of the basics of information science and how a search engine works) can have a dramatic impact on user satisfaction and successful use of the systems. In other words, training the users on how the systems work (not just on how to use the system) and designing systems to work with an understanding of user information behavior may have a dramatic impact on their successful use.

Grant, S., & Schymik, G. (2014). Using work system theory to explain enterprise search dissatisfaction. In *Proceedings of the Information Systems Educators Conference ISSN* (Vol. 2167, p. 1435).

LLMs and AIGC

- **Technology**
- There is virtually no published research on enterprise search implementation and management, so how do vendors know what they need to deliver?
- The default enterprise search application (95%+ market share) is Microsoft Search. How will LLM applications replace/sit alongside the incumbent application?
- **Content**
- Can you validate the information you have found via AI-supported search ?
- Will it stand up to client/customer/governance scruitiny?
- **Training**
- What are the cost/support/training (employees and test collections) implications?
- If employees are having to use workarounds to track down information what is the potential risk to themselves and their organisation?

Two big questions

Can you find it?

Can you trust it?


Search manager