

Bridging the gap between search and IR

Martin White

Managing Director, Intranet Focus Ltd.

martin.white@intranetfocus.com

ISKO



Search
Browse
Monitor
Discovery



Information
Governance
Technology
Intranets





JRC SCIENTIFIC AND POLICY REPORTS

Enterprise Search in the European Union: A Techno-Economic Analysis

Authors: Martin White, Stavri G Nikolov
Editors: Shara Monteleone, Ramon Compañó,
Ioannis Magiros

2013

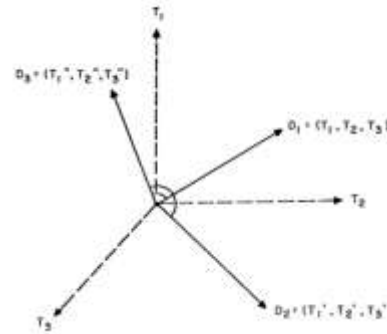


Report EUR JRC86002-01

Joint
Research
Centre

In the beginning.....

Fig. 1. Vector representation of document space.



Gerard Salton 1965

sgi

Big Data ...
and the Next Wave of *InfraStress*

John R. Mashey
Chief Scientist, SGI



Technology Waves:
NOT technology for technology's sake
IT'S WHAT YOU DO WITH IT
But if you don't understand the trends
IT'S WHAT IT WILL DO TO YOU



John Mashey 1997

www.intranetfocus.com



Charles Bourne 1963

How critical is finding the right information to your organisation's business goals and success?

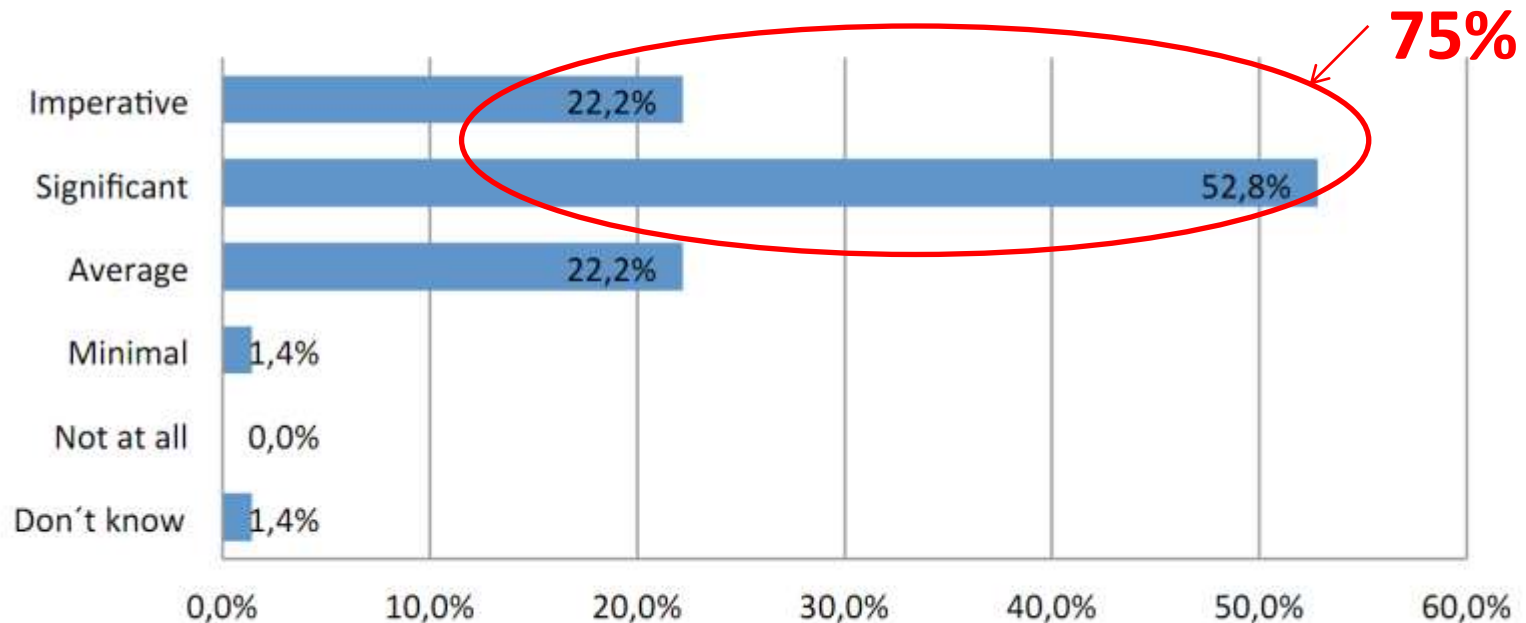


Figure 3 How critical are finding the right information to your organisation's business goals and success?

Enterprise Search and Findability Study 2012 www.findwise.com



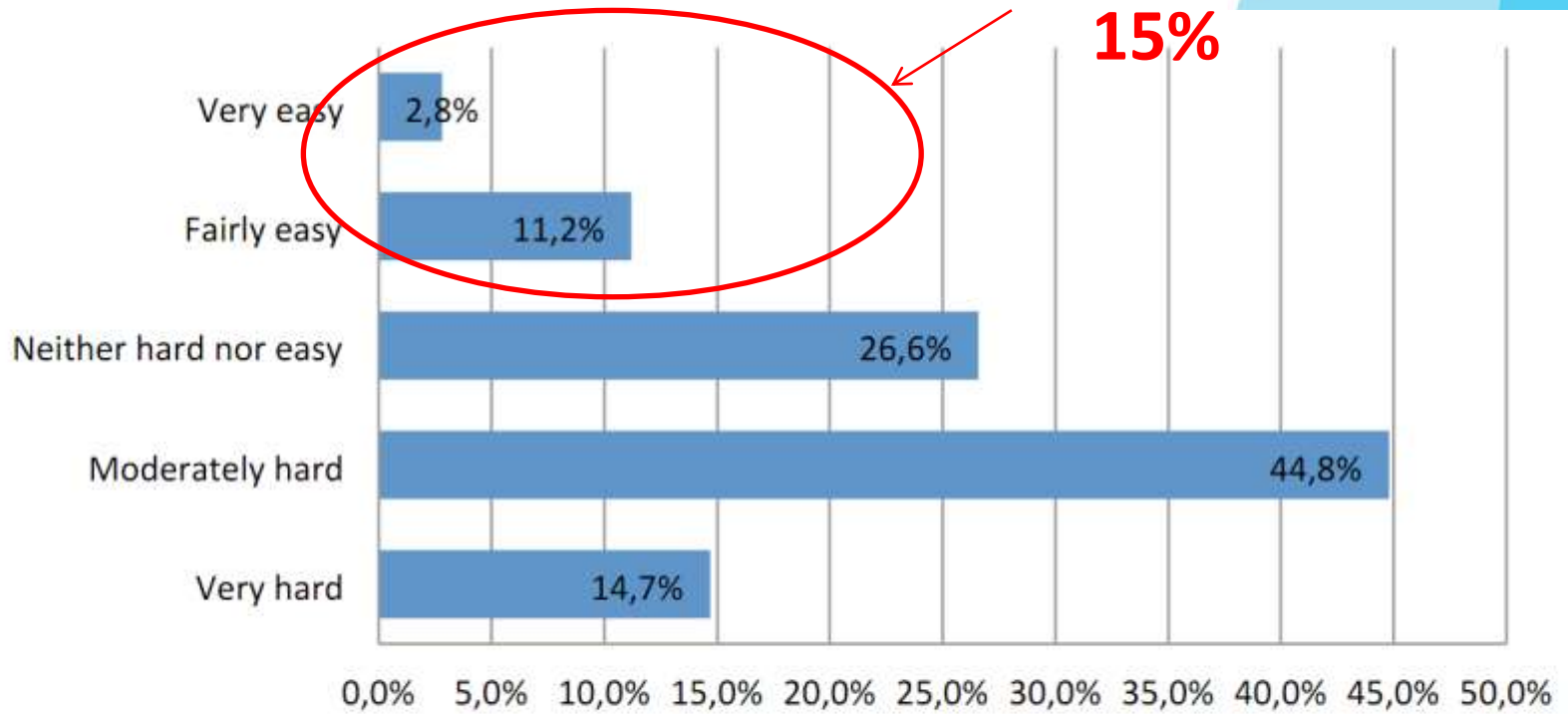


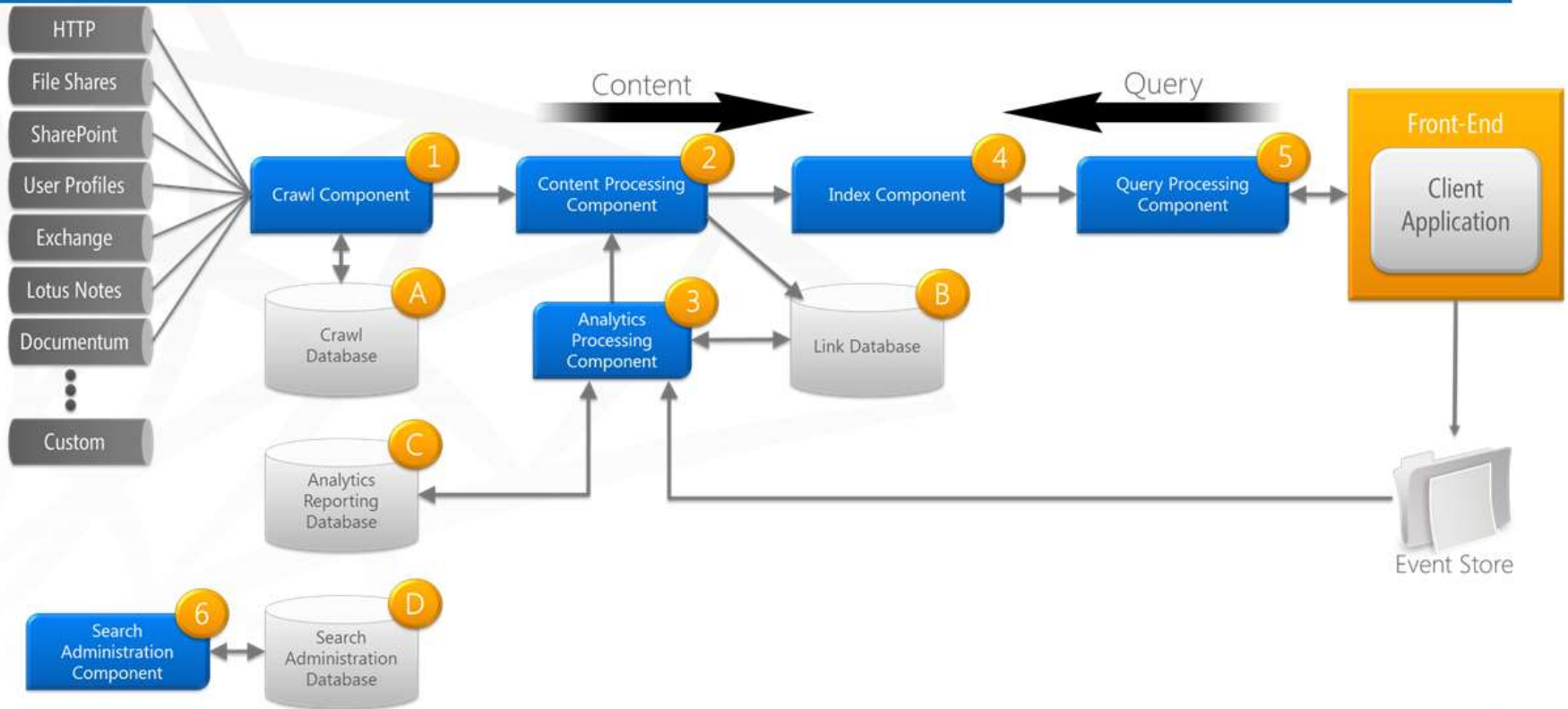
Figure 1 Is it easy to find the right information within your organisation today?

Enterprise Search and Findability Study 2012 www.findwise.com



Search is complex

Logical Architecture



The reality!

How many employees (full time equivalents) are working on administering/developing/implementing search within your organisation?

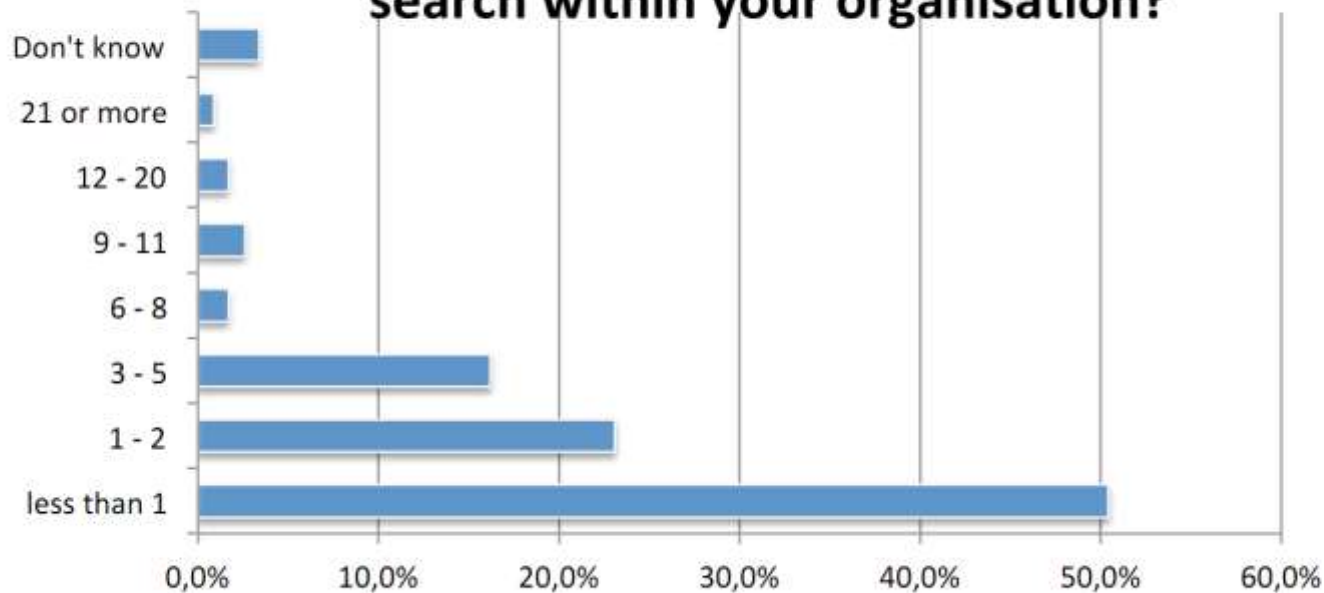
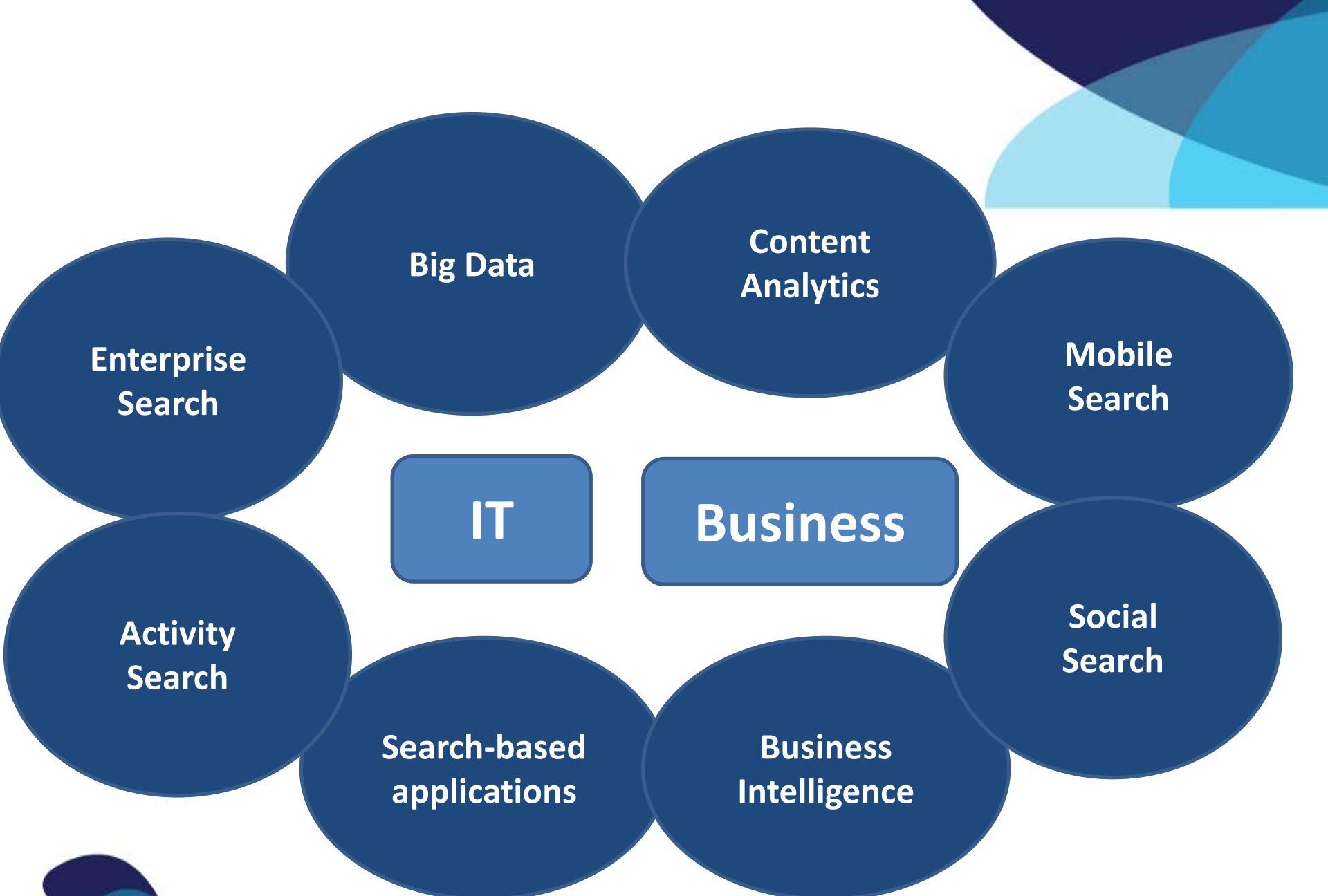


Figure 14 How many employees (full time equivalents) are working on administering/developing/implementing search within your organisation?





The search team!

Images removed for
copyright reason



Techniques and Applications for Sentiment Analysis

**The main applications and challenges
of one of the hottest research areas in
computer science.**

BY RONEN FELDMAN

Over 7,000 articles have been written on the topic. Hundreds of startups are developing sentiment analysis solutions and major statistical packages such as SAS and SPSS include dedicated sentiment analysis modules. There is a huge explosion today of ‘sentiments’





Searching for: sentiment analysis ([start a new search](#))
Found **5,431** within *The ACM Guide to Computing Literature*

Limit your search to [Publications from ACM](#)

REFINE YOUR SEARCH Search Results

▼ Refine by Keywords



Searching for: information retrieval ([start a new search](#))
Found **159,032** within *The ACM Guide to Computing Literature*

Limit your search to [Publications from ACM](#)

REFINE YOUR SEARCH Search Results

▼ Refine by Keywords

 1 A D

Searching for: entity extraction ([start a new search](#))
Found **19,744** within *The ACM Guide to Computing Literature*

Limit your search to [Publications from ACM](#)

REFINE YOUR SEARCH Search Results

▼ Refine by Keywords



Searching for: semantic search ([start a new search](#))
Found **81,080** within *The ACM Guide to Computing Literature*

Limit your search to [Publications from ACM](#)

REFINE YOUR SEARCH Search Results

▼ Refine by Keywords

The current situation

- There are no 3yr undergraduate courses in information retrieval taught anywhere in the world
- There are many universities that have IR research departments and offer MSc and PhD opportunities
- Most courses in information science teach IR as a module and in general do not cover enterprise search management. E.g.
 - Business planning, resourcing and management
 - Search application specification, selection and implementation
 - Search tuning on diverse, heterogeneous content
- Very difficult to gain access to enterprise test collections



Big Data

- However every search application vendor is now positioning themselves to be in the Big Data market
- Universities are rushing to provide Data Science courses, initially at Masters level but now at undergraduate level
- But Big Data without enterprise search does not work

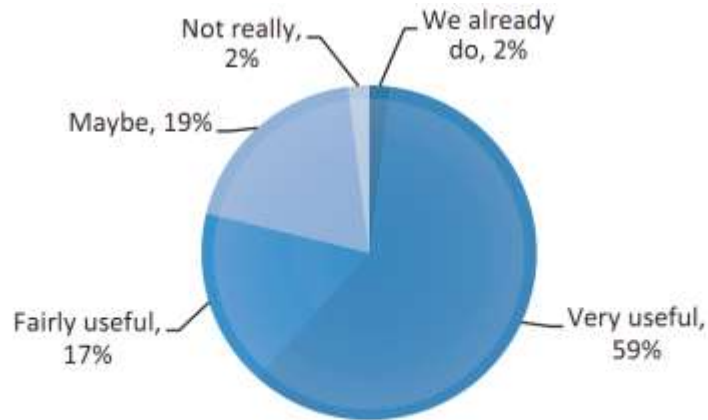
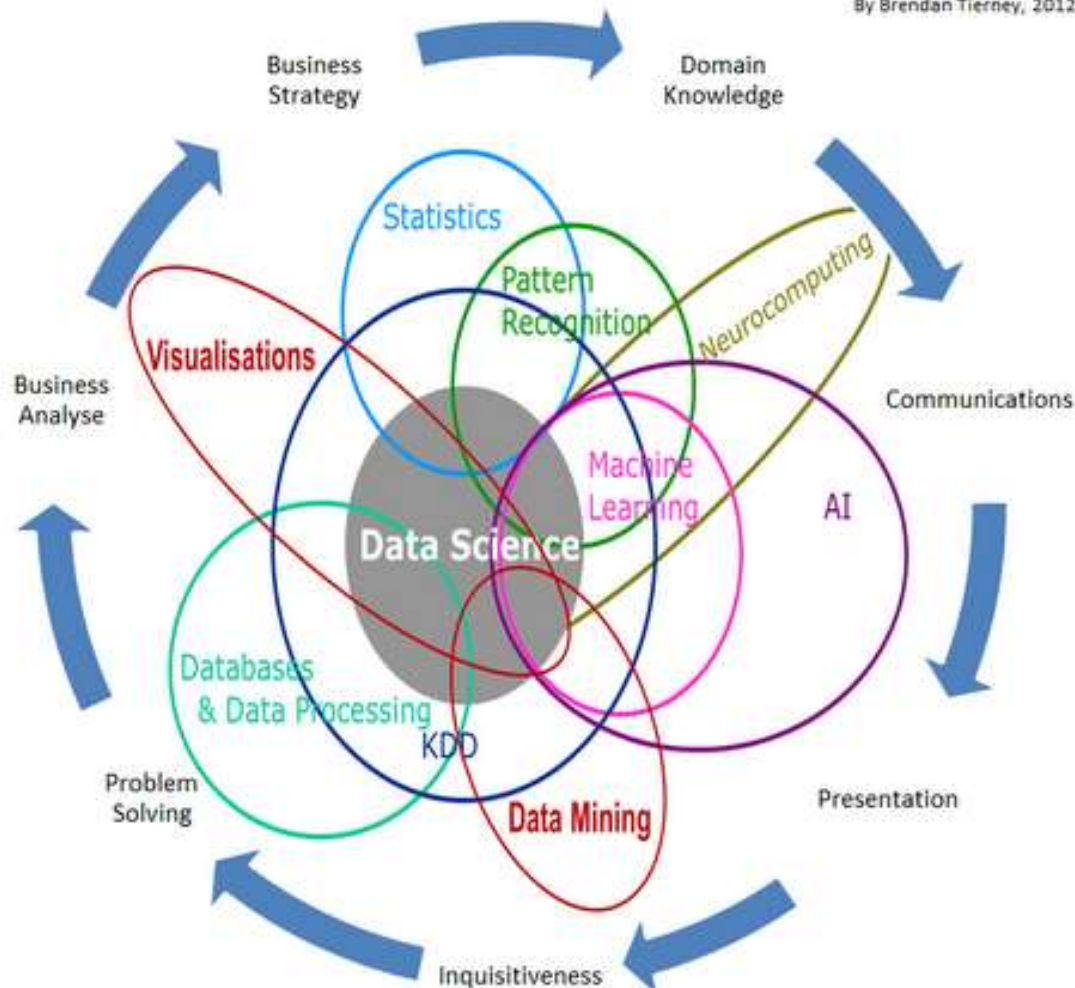


Figure 7: Would it be useful if you could link structured and unstructured data sets for analysis? (N=285)

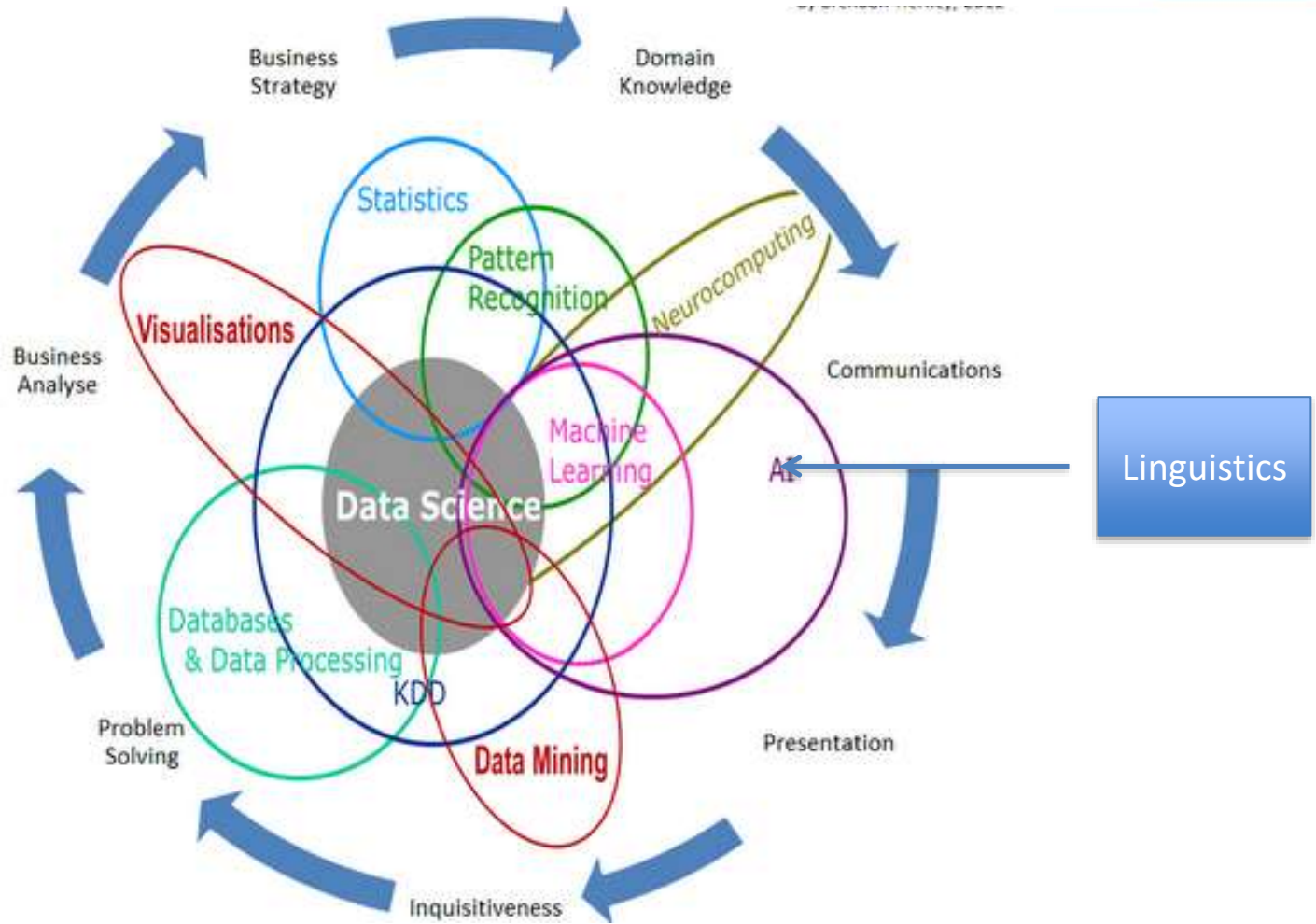


Data Science Is Multidisciplinary

By Brendan Tierney, 2012



Information Science is Multidisciplinary



An enterprise perspective

- No courses available on enterprise search management
- No awareness of developments in search technology that could eventually appear in commercial applications or be incorporated into open source applications
- Most courses in information science teach IR as a module and in general do not cover enterprise search management.
- There are around 100+ books on IR but only one on enterprise search
- There are no mechanisms to locate and obtain academic research papers

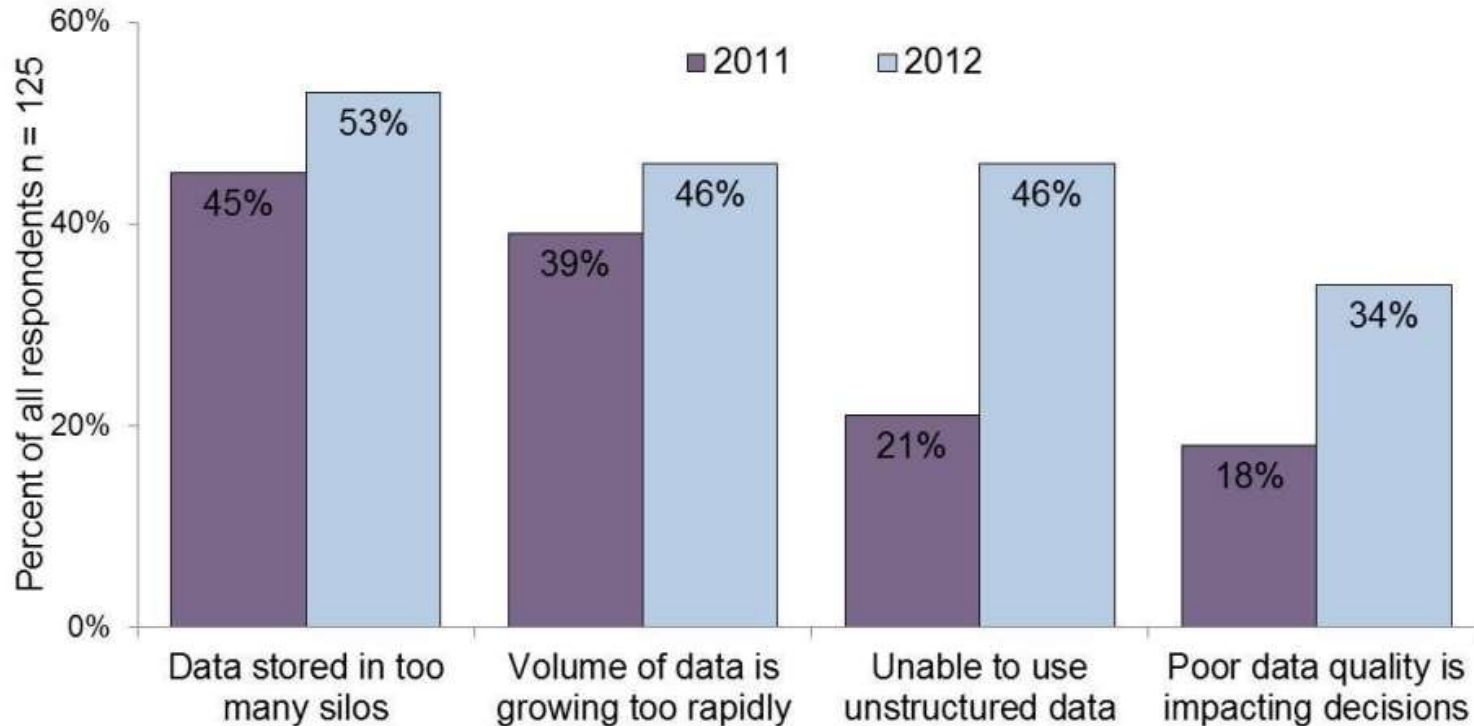


The implications

- Search vendors and search systems integrators cannot find the staff they need to support development and customer support
- Open source developers find it difficult to keep up with IR developments
- Organisations cannot find the staff they need to create search support teams
- HR departments do not have an appreciation of skills, salaries and career development for search specialists
- HR agencies also have little idea of 'search'



Figure I: Business Pressures Driving Data Investment



Source: Aberdeen Group, January 2013



Actions

- Universities need to develop enterprise search/big data courses at undergraduate level
- Enterprises need to make test collections available to the IR community
- Enterprises need to be open to working with IR research teams
- Publishers and academics should look at how IR research can be made more accessible
- There is probably scope for an inter-disciplinary journal on ES
- IR and ES conferences should be more accessible to the two communities



The Search Circle



Questions and comments

