

META-data in search & BehaviorTracking + Feedback Mechanism

Anders Hyldahl

Chief Software Architect, CTO

Mondosoft



Mondosoft

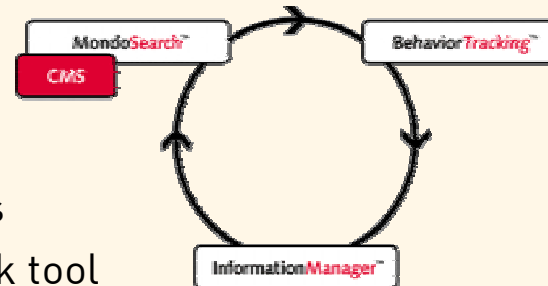
About Mondosoft and MondoSearch

History:

Mondosoft was founded in 1998 in Denmark
1800+ customers in 2005 worldwide

Products:

MondoSearch – Corporate Search Engine
BehaviorTracking – Search Behavior Analytics
InformationManager – Search Behavior Feedback tool



Locations:

US Headquarter – Palo Alto, CA
EU Headquarter – Copenhagen, Denmark



Mondosoft

We are standing on other's shoulders

Ptolemy 2nd founded the library in Alexandria, Egypt, in 283 B.C.
based on ideas by Aristotle on how centers for science
and knowledge should be. In periods more than 500.000
books was collected in the library

Zenodotus (born c. 325 B.C.) – the first librarian in Alexandria
Alphabetized texts

Callimachus (c. 305-240 B.C.)– a librarian
Invented bibliography and wrote the first catalogue on texts

Philetas (born c. 320 B.C.)
Invented the idea on a dictionary and wrote the first one

In periods more than 100 persons worked fulltime in the library in Alexandria. The
library was destroyed in ancient time, maybe burned down by Caesar in 48 B.C.,
maybe in another way...

Elements of Search technology today

- Text-mining / database engines
- Ranking algorithms
- Clustering / Categorization
- Synonyms, stemming, spelling correction etc.

Text-mining

Ranking

Clustering

Thesaurus



The library in
Alexandria today



Mondosoft

The way META-data is used today

Sample

Search for 'Elephant' on global web search engine

Results

- More than 10 million documents found
- We use other information than the ranking to choose the document to view
- We are helped if:
 - documents are grouped by categories
 - documents have informative titles, descriptions or other descriptive information
 - documents are sorted in ways that suites the type of information we are looking for



Mondosoft

More information than the documents it self

META-data is information about information

Common document META-data: Title, Creation date, Document size and URL

Other document META-data: Author of the document, Abstract, Editorial facts, Publish date, Publisher, Technical information, Cluster etc.

Tasks:

- 1) Identifying and collecting META-data
- 2) Define ways and tools for using META-data in search



Mondosoft

Excuse: From the history of mathematics

- 1) Euclid, Archimedes and others (400 B.C. – 100 A.C.) used symbols as variables in description of geometry. That introduced very advanced mathematics with base in geometry
- 2) By introduction of the coordinate system by Descartes and others the geometry took a big step into coordinate geometry
- 3) The mathematical Analysis was rising when Newton, Leibniz, Euler, Lagrange introduced symbols and notations for solving problems that has been known for long time
- 4) The ancient and not very organized formal logic founded by Aristotle got a new nature by the methods introduced by Boole in 1847 and other mathematicians

Having a formal language on a given area develops the area it self



Mondosoft

The effect of having a formal language

Samples of formal languages:

- 1) Programming languages like Java, C++ or Pascal
- 2) SQL (Structured Query Language) for accessing a database

In text retrieval:

A formal search query language can be used to specify how to use META-data for filtering, grouping, sorting/ranking, result-set content and much more

A formal search query language will enhance and develop the way search and META-data is used and works



Mondosoft

Mondosoft Query Language

- 1) MQL (Mondosoft Query Language) is a formal language to formulate searches and use of META-data
- 2) MQL is meant to be used by the system integrators for implementing search functionality in a corporate environment
- 3) MQL enhance the use of META-data in search
- 4) MQL will develop during time like other formal languages

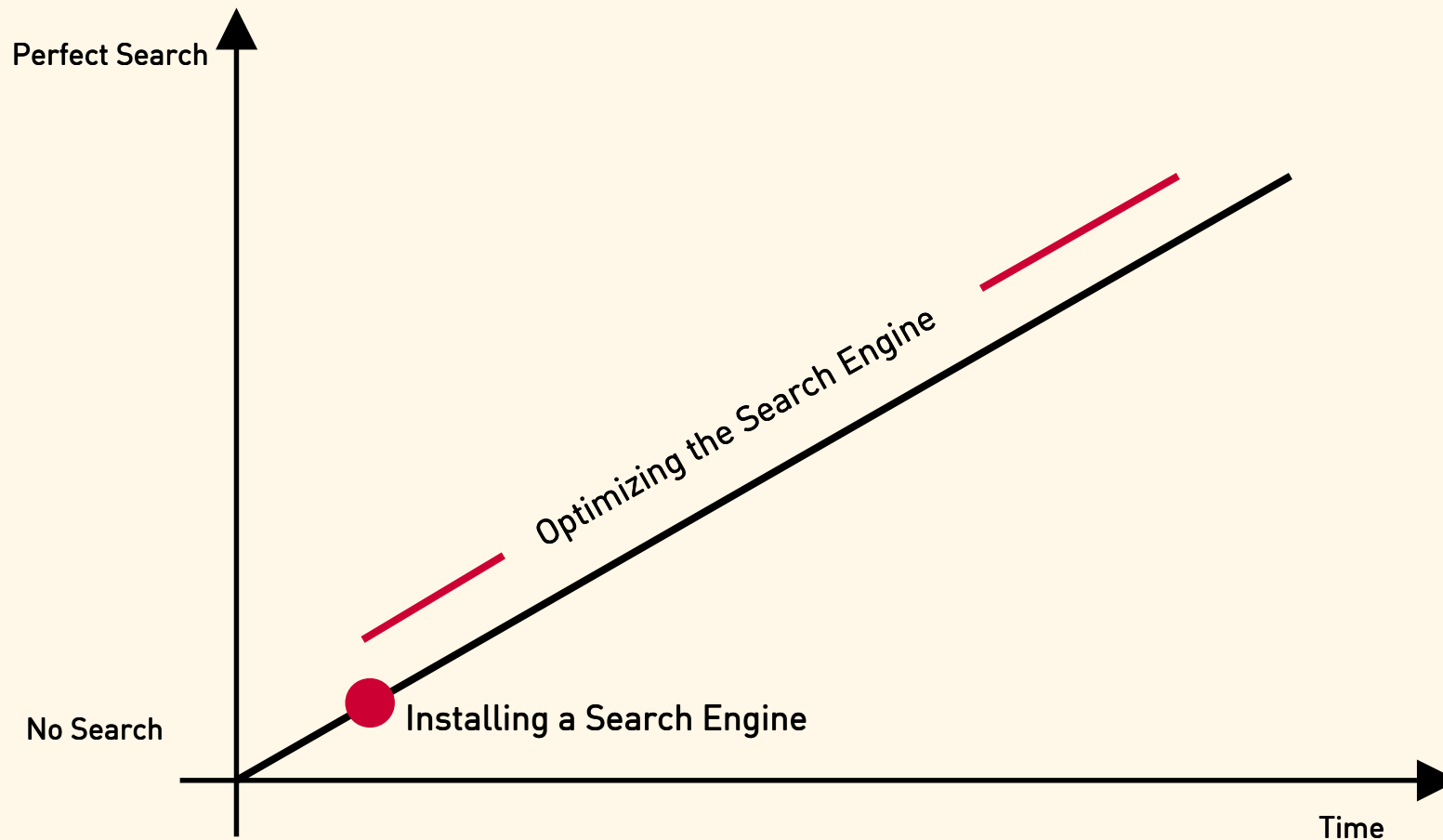
Sample

```
OPTIONS SearchQuery='Elephant', SearchExcludeQuery='Africa Animal',  
StemmingOn, HighlightOn  
FILTERS (MimeType='text/html' OR MimeType='application/pdf') AND  
(Language='en' OR Language='de') AND DocumentDate>2002-12-31  
ATTRIBUTES HtmlTitle, Author, DocumentDate, Language, MimeType,  
HasImage, HasVideo, HasSound  
LIMITS MaxResults=20
```



Mondosoft

Implementing a Corporate Search Engine

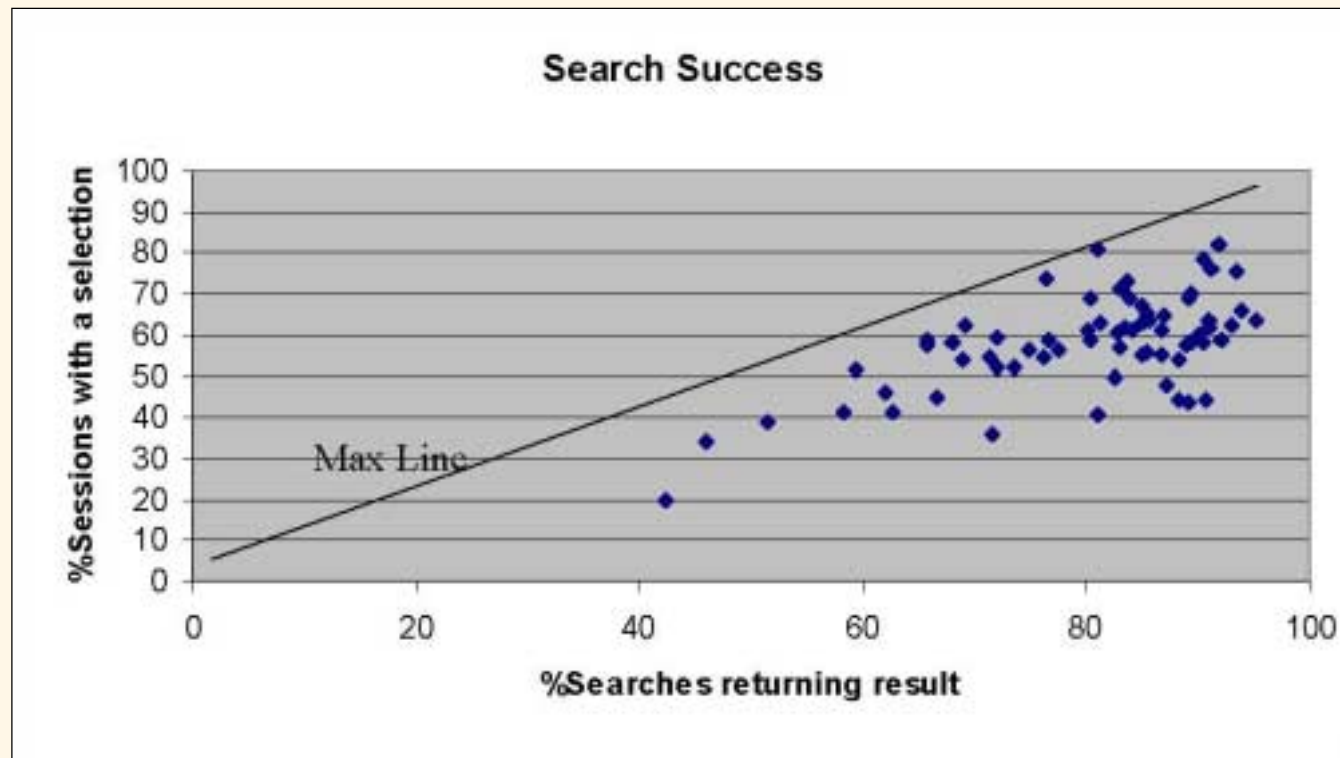


Mondosoft

Search performance

- North indicates a better search implementation
- East indicates relevant content delivery

SSI (Site Success Index) = Distance from North-East corner

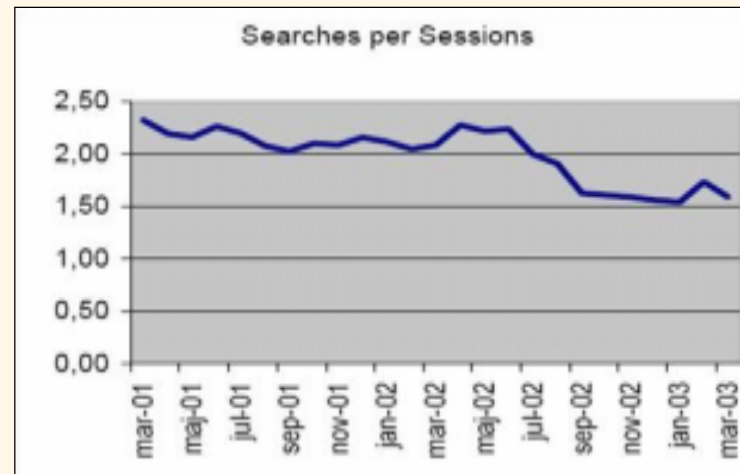
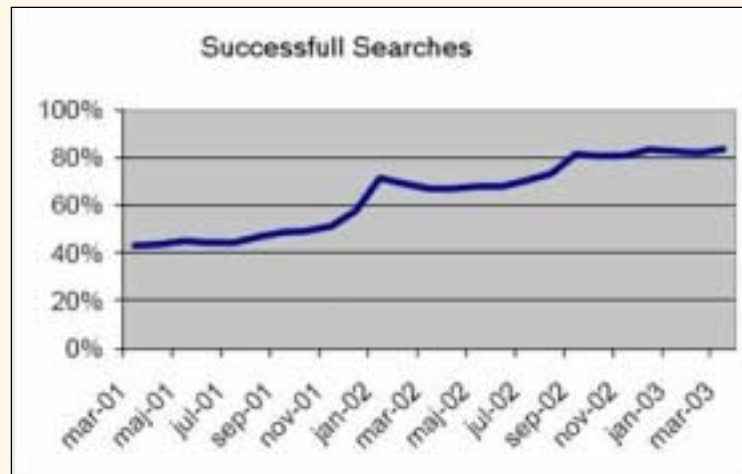


Mondosoft

Behavior Tracking & Feedback optimization

By using Behavior Tracking and Feedback Corporate Search Engines can be optimized

Sample: Peoples Bank (www.peoples.com) uses MondoSearch, BehaviorTracking & InformationManager and over time they have improved the search quality



Mondosoft



Mondosoft