I was going to talk about
“Information Retrieval in extreme contexts”

But I will talk about “Aggregation”

Mounia Lalmas
mounia@acm.org
Some sort of outline

- Document retrieval, information retrieval, web search
- Focused retrieval
- Aggregation
- Current work - Digest page
Classical document retrieval

1. doc1
2. doc2
3. doc3

Document corpus

Retrieval System

Ranked Documents
Classical document retrieval process

Query Representation

Retrieval Function

Ranked documents

Index

Representation Function

Documents

Query

Representation Function

Document Representation
Information retrieval process

Query

Representation Function

Query representation

Retrieval Function

Index

Documents

Representation Function

Object representation

Results

Task
Context
Interface
Interaction
Multimodality
Structure
Media
Passage
Mobile
Language

Peter Ingwersen & Kalervo Järvelin, The Turn, 2005
Information retrieval and web search
Question:

When was Madonna born?

Possible answers:

1. Madonna was born Madonna Louise Ciccone in Bay City, Michigan on August 16, 1958. Madonna was the third of eight children born to Madonna Fortin and Sylvie Ciccone.
2. Madonna (read more) 5,958,624 plays scrabbled on Last.fm Madonna Ritchie (born Madonna Louise Ciccone on August 16, 1958 in Bay City, Michigan, USA) is an iconic pop singer, dancer, song writer, producer, author and actress.
3. Madonna Madonna was born Madonna Louise Veronica Ciccone in Bay City, Michigan, August 16, 1958, and grew up in a working-class household in Detroit.
4. Madonna was born Madonna Louise Ciccone in Bay City, Michigan, in 1958.
5. knowledgebroker -- Madonna was born as Madonna Louise Ciccone in Bay City, Michigan on August 16th, 1958.
7. Madonna, born Madonna Louise Veronica Ciccone in 1958, was the third of eight children.
9. Originally performed by Madonna (born Madonna Louise Ciccone 8/16/58 in Bay City Michigan)
Focused Retrieval - XML Element Retrieval

Search Result

1: (0.247) Scalable Feature Mining for Sequential Data
Neal Lesh Mitsubishi Electric Research Lab Mohammed J. Zaki Rensselaer Polytechnic Institute Mitsunori Ogihara University of Rochester
Result path: /article[1]/bdy[4]/sec[5]

2: (0.204) Probability and Agents
Marco G. Valtorta University of South Carolina mgv@cse.sc.edu Michael N. Huhns University of South Carolina huhns@sc.edu
Result path: /article[1]/bdy[4]/sec[3]

3: (0.176) Combining Image Compression and Classification Using Vector Quantization
Karen L. Oehler Member IEEE Robert M. Gray Fellow IEEE

4: (0.175) Text-Learning and Related Intelligent Agents: A Survey
Durja Miladenic J. Stefan Institute

5: (0.175) Detecting Faces in Images: A Survey
Ming-Hsuan Yang Member IEEE David J. Kriegman Senior Member IEEE Narendra Ahuja Fellow IEEE
Result path: /article[1]/bdy[4]/sec[2]/ss1[9]/ss2[10]
Aggregated (virtual) documents
Aggregated (virtual) documents

- Web search
  - Clustering (Clusty)
  - Summarisation (WebInEssence)

- News domains
  - Topic detection & tracking (TREC track)
Aggregated views
Aggregated views

1. Visit London - London's official website - London Tourist ...
   Visit London with ease using London Tourist, Entertainment and Travel Guides. ... those 
   tiaras and feather boas – the world's campest musical is back in London! ... 
   http://www.visitlondon.com/

2. LondonTown.com | The Number One Internet Site for London England ... 
   Sightseeing, attraction and other tourist information. Theater tickets, London maps and more. 
   ... Christmas Lights in London 2007 ... 
   http://www.londontown.com/

3. London Entertainment Guide from The Evening Standard | This is London 
   ... tickets booking for London covering Film, Restaurants, ... London 'safest place' ... This is 
   London Magazine Ltd. Daily Mail Mail on Sunday Travel Mail ... 
   http://www.thisislondon.co.uk/

4. London - Wikipedia, the free encyclopedia 
   An important settlement for two millennia, London is one of the world's leading ... London is a 
   major tourist destination with four world heritage sites, several ... 

5. Mayor of London, the London Assembly and the Greater London Authority 
   The official website for the Mayor of London, the London Assembly and the Greater London 
   Authority ... London Assembly members, Richard Barnes and Tony Arbour, ... 
   http://www.london.gov.uk/

6. Time Out London - Discover Things To Do in London Including ... 
   Highlights this month's events in London and provides details of accommodation, ... London 
   Jazz Festival ... London ice-skating rinks. Start practising your ... 
   http://www.timeout.com/london/

7. BBC - London - London Homepage
Current work

- Digest page
  - Yahoo! Wikipedia
  - Clustering (Carrot²)
  - Multi-document summarization

- Ranking of digest pages

- Aggregated digest page
  - Query context
Current work - ranked list of digest pages
Current work - presenting a digest page
Current work - context of query
Aggregated “answers”

Why?

- reduce information overload (is there one?)
- increase result space (is this needed?)
- provide/capture context/focus (country, language, device, literacy, …)
- presentation and interface (expectation for Google type interface very high! Is it the case?)
- cross/multi-media content (refer to StoryBank)
- cross/multi-lingual content (is there a need?)

Is the above too complicated? Not needed? Not feasible? In 10 years?