

2nd CHORUS Conference and 3d Workshop on Future of Web Search: Beyond Text

Jussi Karlgren

Report

The Second CHORUS Conference and third Yahoo! Research Workshop on the Future of Web Search was held during April 4-5, 2008, in Granvalira, Andorra to discuss future directions in multi-medial information access and other specialised topics in the near future of retrieval. Attendance was at capacity, with 97 participants from 11 countries and 3 continents.

Identified challenges

Several different research strands were represented at the conference. Several challenges to the field at large and to the specific projects discussed were presented by the speakers. Some themes that recurred were the possibilities of progressing from text retrieval to content-based retrieval of non-textual data (or the rationale for not doing so, based on more refined models of the textual information associated with the data in question), the utilisation of user-contributed content and the encouragement for users to provide more such content and of higher quality, and the description and evaluation of use cases and business models for new services in multi-media retrieval. A tentative generalisation is given by the four following points.

Search Strategies and Session Design

The interaction point between user and system can be designed to conform to a search scenario, a browse scenario, or a recommendation scenario, to name some obvious and well-established access paradigms — and naturally, this is not an exhaustive list. Other types of interaction are possible. In the presentations, several speakers discussed this dimension of variation. Is this a modular issue, not related to underlying system architecture, or does the choice of interaction paradigm have ramifications for engineering the system itself? And does the character of the data, viz. multi-medial data in the case at hand, influence or incur preferences for the choice of possible session designs? There is an entire research field devoted to interaction design, but its methodology and generalisable results are seldom applied directly to the information retrieval field: the methodological gap between the fields of interaction design and information access needs to be bridged in future projects. The CHORUS concertation action has focussed much of its effort into defining use cases for access to multi-media which may be a support framework for projects to provide generalisable results in this respect.

User-contributed data

Many of the presentations discussed user-contributed data. The lowered publication threshold and the true, if not completely symmetric, bi-directional communication technology allows users to contribute content, and structure it at will and on demand. A recurring theme in the projects presented was how to provide a framework which a) encourages and motivates users to contribute to shared information systems and b) provides guidance, quality assurance, and a shared semantic space to contributions, better to build a common body of knowledge. These issues link in with questions of socio-economic and legal aspects of multi-media distribution.

Representation and architecture

Several presentations discussed how to analyse multi-medial information objects to enable search and indexing of them; other presentations used existing proxies for information such as user-contributed comments or tags; yet others used contextual data mined from user behaviour or meta-data harvested from extraneous knowledge sources. This provoked some debate as to whether data in different media are of inherently different type: are textual data primary and more refined where e.g. video data are more raw, in some respect? Can we envision a general representation not tailored to some task or domain? Should we aim to build e.g. a language-based representation for data which haven't been expressed in linguistic terms — or is text a false target for representation? Can we presume to know what the tentative uses of some information object will be? In addition, the question of processing architectures, algorithm optimisation, and system design are likely to have effects on the usefulness, efficiency and effectiveness of resulting systems.

Evaluation

Taking the various challenges together brings the problem of evaluation to the fore. Evaluating information retrieval has traditionally been done using the target notion of topical relevance. How this target for quantitative relevance can be extended to non-topical access scenarios is not obvious — should it be supplanted by more general notions such as user satisfaction or pertinence or should the notion of relevance be enhanced?

Conference Program

The program included 8 project presentations from European projects SEMEDIA, MESH, RUSHES, SAPIR, AIM@SHAPE, VITALAS, TRIPOD, PHAROS all gave brief introductions to their project starting points, aims, and results; industrial participants Telefonica, Motorola, Polar Rose, Exalead gave a view of current trends in commercially funded technology development; public and industrial research laboratories Fundacio Barcelona Media, CNR, A9, Yahoo! Research Barcelona, and Yahoo! Research Berkeley gave a foundation of research results to build on. In addition the conference began with an invited presentation by *Wessel Kraaij* of TNO, Amsterdam, giving an overview of the TRECVID video retrieval evaluation and concluded by a presentation *Usama Fayyad*, Yahoo Vice President of Research and Strategic Data Solutions, giving an outline of the commercial basis for the near future of the search industry. The first day included a session for informal discussions with several demonstrations of presented systems and a hands-on workshop by *Ernesto Arroyo* of Fundació Barcelona Media who demonstrated mouse tracking technology for tracking user behaviour.

Wessel Kraaij's initial presentation provided a broad perspective of how and why video retrieval differs and is similar to evaluation of text retrieval, and of the state-of-the-art of current video retrieval technology. Among the several factors brought up of general relevance is the observation that currently only few projects even experiment with content analysis of the video signal itself, mostly limiting themselves to metadata analysis and extraction of linguistic data from the audio stream, and that the publishing threshold for video is higher and the sources more diverse in many respects, which gives no search engine alone coverage over a significant proportion of video sources and data. This differs importantly from the text search field, where several contenders have a fairly large cross-source coverage.

On interaction design

Addressing one of the challenges from Wessel Kraaij's presentation, *Xavier Amatriain* of Telefónica Investigación i Desarrollo discussed the interface between search and recommendation, arguing that in multi-media information access we will naturally be moving from search to recommendation. This presentation neatly brought together several of the discussion strands of the workshop: does lean-back interaction and system initiated interaction necessarily involve recommendation rather than search? Are the data under consideration inherently more suitable for recommendation? Will the recommendation paradigm complement the searching paradigm or merge with it?

An example case study was given by *Arjen P de Vries* of CWI, presenting the VITALAS project, which is based on a use case analysis and works simultaneously with both textual and example-based retrieval of images. VITALAS interfaces support both recommendation and search, and the research questions under consideration were on how to use both personalised recommendation and user expressions of information need as basis e.g. for cross-medial relevance feedback.

On personalisation and other contextual factors

Adding personalisation to recommendation and search will imply opportunistic action on the part of the system, opening for system-initiated information provision interleaved in other sessions. As a signal of sessional history and user-related data having impact, *Usama Fayyad* presented convincing business proof of personalised search effectiveness: if users can be tracked and advertisements served based on their past search history, the impact of those advertisements on users is considerably larger than for non-targeted advertisements. This, of course, is a factor not limited to advertisements but to any information delivery interface.

On the topic on how search might be extended to other platforms, *Ben Bratu* of Motorola presented a number of industrial challenges to providing useful personalised mobile search solutions — how personalisation can be made transparent and tangible, and how the information needs can be recommended to users based on actions of others. One conclusion made from previous projects at Motorola was that recommending queries rather than content itself for user approval was more suitable for constrained information access situations, to keep interaction simpler and shorter: the challenge is not to build a new search engine for every new situation and session configuration, but to build a sustainable and cross-platform recommendation framework. In his presentation, *José Manuel Cantera Fonseca* of Telefónica Investigación i Desarrollo described how mobile search can be made more responsive to the constraints of the mobile platform, without explicit effort being made on part of the user; ranking results and determining if target documents in the database are useful for mobile users using a representation of suitability for mobile presentation.

How a representation can be built from data and user behaviour for practical usage was shown by *Julien Law To* of Exalead, presenting new indexing mechanisms and *Oscar Celma* of Universitat Pompeu Fabra, presenting the PHAROS project, both exemplified how opaque audio-visual data container streams can be analysed into internal structure, both by content analysis and by informed analysis provided by the originator of the data: the examples given by the PHAROS project was how video streams can be segmented using information from the audio track and by Exalead how access to news clips can be directed to the time where some sought index term is pronounced, using language as the index term and extracting salient entities from the analysed audio track. In approaches based on modelling usage-based needs, *Marco Pellegrini* of CNR presented the VISTO visual storyboard which combined content analysis with a timeline-styled display of video sequences for

browsing video clips and *Pedro Concejero*, of Telefónica Investigación i Desarrollo, presenting the MESH project, showed the project plan for a personalised multimedia summary of news items — based on both content analysis and matched to user profiles.

On the question of representation

Jan Erik Solem of Polar Rose presented a combination of content-based extraction of salient items from images with social network based annotation applied to recognition of individuals in photographs. Polar Rose is currently introducing its system into commercial use for the general public. The approach hinges crucially on creation of a level of representation which would be appropriate for general usage: not only what features can be recognised but what items can be named and communicated adequately to image viewers. What level of representation might be considered to be natural and general enough to be used outside immediate focussed task scenarios? An example of how to use existing textual or linguistic representation schemes for annotation of non-linguistic objects was given by *Francesco Robbiani* of CNR, presenting the AIM@SHAPE project, showed how three-dimensional representation and acquisition, how to specify and to acquire a nomenclature using existing ontologies of 3D objects mapped to shape-represented 3D objects. Linking raw geometry to structure to semantics. *Pedro Concejero* of Telefónica Investigación i Desarrollo, presenting the RUSHES project, also discussed the need of an open set of mid-level metadata to enable high-level semantic queries — using content and structure descriptors of an appropriate level of abstraction.

Roelof van Zwol of Yahoo! Research Barcelona presented how the SEMEDIA project is investigating user behaviour in annotating images and how images on the Flickr photo sharing service are shared and viewed by web users. As many other human behavioural statistics, image viewing can be matched to a power distribution, where a few images account for much of the interest and a huge proportion of images are viewed and annotated by few users. Mining these data afford both a challenge and an opportunity to better understand the character of how tag annotations can be understood. This same discussion was addressed again by *Vanessa Murdock*, also of Yahoo! Research Barcelona, who in her presentation on image tagging pointed to what she termed a *Language Disconnect* between the language used by the potential viewers of images and the language used by the original uploaders when annotating them. Queries and tags are done with different aspects of the original image in mind. How to bridge that gap? This points directly to the challenge of representation.

On enriching the representation

What other data besides immediate indications of topic might then be useful for enriching information access to multimedial data?

Xin Fan of Sheffield University, presenting results from the TRIPOD project, and *Mor Naaman* of Yahoo! Research Berkeley both demonstrated the utility of using geographical and positional information, both to improve access to data and to extract information from it. Arguably, geographical information, both absolute positional information and indirect, landmark-oriented information is cognitively special and central to our understanding of events, objects, and action. Making use of this information is obviously both effective and useful, but quality control of the contributed information is central to the usefulness of the resulting knowledge. Since the long tail of user-contributed information is for all practical purposes endless — there are unending possibilities for variation in this type of data — a conservative approach to deduction from the data is advisable. *Omar Alonso* of A9 added a further dimension to the spatial data, by noting the importance of temporal data

in understanding information objects: time is central, salient, and frequent in texts (as well as other media) – which is being explored in several different experimental interfaces at present, and mentioned in the search log analysis performed with the VITALAS project presented by Arjen P. de Vries. Exploitation of temporal data has not yet been settled — geographical data can be used to plot items on e.g. a map, but temporal data are not necessarily as plottable and absolute.

Another example was given by *Peter Mika* of Yahoo! Research Barcelona, who gave examples of how existing information on the net can be leveraged to provide more enriched data representations in the form of metadata-driven *microsearch*. Instead of requiring formal representation schemes for enriched data, much of this data can be included by examining existing resources this sort of enriched information automatically learnt and generalised from specific data found in existing sources. An example application demonstrated at the conference is *people search* where data about people sought for were collated from multiple sources, including e.g. geographical data.

On algorithms and optimisation

In any case, introducing further data of whatever type adds complexity to the challenge of representing information objects appropriately. *Mihai Surdeanu* and *Aristides Gionis*, both from Yahoo! Research Barcelona, demonstrated how information access using extraneous information from the immediate surface representation of documents themselves can improve specific results — if feature combinations are done right. This is a representational challenge: added information, whether information about entities mentioned in a text or contextual information about links to other documents, can result in noise rather than improvement if algorithms for utilisation are designed sub-optimally. Related to the design of algorithms is the design of architectures, and *Pavel Zezula* of Masaryk university, presenting the SAPIR project and its MUFIN framework for general-purpose retrieval, discussed the tradeoff between scalability and determinism inherent in going from a single-server architecture to a distributed architecture or an overlay network and stressed the need for system efficiency to provide useful infrastructures for whatever type of indexes are at hand.

Conclusions

The challenges outlined in the introductory section will be discussed further at coming events. Some points can be given as a partial conclusion of the conference discussions.

- Progression from text search systems to multimedia proceeds; systems are being deployed for industrial and public use at present; challenges remain both on practical and conceptual level, and the design of data representation, algorithms to access them, and architectures to house them within cannot proceed without sensitivity to each other.
- Service design, interaction design and use case analyses are a crucial challenge for projects: how can the technological advances be refined into services which either feed into existing business models or provide new business models with a realistic entry threshold?
- User generated content and bidirectional publishing flows will, if they provide both the right conceptual infrastructure and the right motivational framework, be a valuable resource for information access applications. How to ensure quality and encourage user participation are design questions both on the level of system design and on the level of data representation.
- Evaluation of multimedia systems needs to be sensitive to all the specific challenges in play: content representation, use case and session design, and further lowered publication thresholds. Evaluation schemes have been the backbone of text retrieval research for the past decades: the field of multimedia retrieval should take care not to diverge from that tradition for reasons of convenience only – the target concept of relevance may not need to be replaced, but will need either extension or deconstruction to carry over to new use cases, new scenarios, and new types of media.