Relevance is a key concept in information retrieval (IR); the main task of the search system is to bring forward items (records) that are relevant answers to the search query.

Relevance is, however, a complex phenomenon. Saracevic (2007) presents a thorough analysis of the research and scholarship that has been carried out on the phenomenon and related aspects. There are two main approaches: system based relevance and user based relevance. The systems viewpoint emphasizes the [automatic] processing of information objects and the matching with queries in an algorithmic way. The user viewpoint, on the other hand, considers the system as given [like a black box] and the user is recognized as an actor that has many cognitive and social dimensions, as well as interactions with the system.

[based on Saracevic (2003, p. 1925)]

We shall use the term System Based Relevance for the relevance manifestations from the systems viewpoint (encompassing system or algorithmic relevance, and topical or subject relevance). We shall use the term User Based Relevance for manifestations from the user viewpoint (encompassing cognitive relevance, situational relevance and affective relevance).

Cross-domain search services bring together material from two or more of cultural heritage domains, such as libraries, archives, audiovisual collections, museums and sometimes «sites and monuments». The contents – the database records – are of different formats and structures which poses special challenges for the search services. To evaluate the relevance of hits is the first challenge; to deliver response sets with high precision is the second challenge.

Traditional information retrieval techniques are not well suited for this new situation. Semantic techniques as exemplified by the Europeana Data Model are put forward as a promising approach.

The paper discusses approaches to deal with the challenges. Semantic techniques as exemplified by the Europeana Data Model are put forward as a promising approach.
Borlund (2003) in her treatment of relevance has a more specific focus: she presents an overview of the multidimensional and dynamic nature of relevance in the context of evaluation of interactive information retrieval systems.

She arrives at the conclusion that situational relevance fulfills the necessary requirement to encompass both the multidimensional and dynamic aspects of relevance. In her definition: “situational relevance is user-centered, empirically based and realistic as well as potentially dynamic type of relevance. Situational relevance expresses the relationship between the user’s perception of usefulness of a retrieved information object, and a specific work task situation.”

Saracevic (2007, p. 1930) raises the question of primacy: “Does topical relevance underlie all others? Do all other manifestations of relevance follow from topical relevance and does it have primacy among relevance manifestations?” Different schools of thought give different answers. In this paper we assume that topical relevance (i.e. system based relevance) underlie, and is necessary for, user based relevance.

The traditional performance measurement precision (the number of relevant hits as a proportion of the total number of hits) is a system based measure of relevance, mostly topical or subject relevance. Low precision, i.e. a large number of non-relevant hits, can easily downgrade the total response to be felt less relevant.

We shall discuss some special issues and challenges for cultural heritage search services with respect to relevance and precision.

To calculate topical (or subject) relevance is a more difficult task in cross-domain services since the material (database) is heterogeneous in terms of record structure and in the semantics of structural elements.

To consider situational relevance we discuss two kinds of audiences – the interested public and the researcher or specialist. Both have an interest in getting relevant results from a search.

The general user can be expected to be more tolerant for near-hits, and less sensitive to precision. Results must not be confusing, but stimulating results are acceptable even if their subject relevance is low.

The specialist can be expected to be more focused and will appreciate precision in the results.

**Precision and the quality of the search service**

**Fielded search**

There are two major traditional precision enhancing techniques: The first is fielded search – directing the search to particular part/s of the records, e.g. the title. This is often combined with boolean operators, such as AND, OR, NOT. Most library search services (online catalogs) offer a wide range of fielded search options under the heading Advanced Search.

If the database is MARC-based there are many fields that can be made searchable and used to specify a query. The Library of Congress is a good example, see [http://catalog.loc.gov/](http://catalog.loc.gov/).

In cross-domain services records of different structure are brought together so fielded search can only be applied to the least common set of field labels, so it is **not possible to offer the same detail in fielded search**. Field labels that are special for one particular domain are sacrificed even though they might add significantly to precision.

In the Europeana portal ([http://www.europeana.eu/portal/index.html](http://www.europeana.eu/portal/index.html) ) Advanced search is limited to the fields; Title, Creator, Date and Subject (the latter initiates a free text search and is not a controlled subject list). In comparison the Gothenburg University Library offers the following fields in Advanced search: Title, Author, Name, Subject, Form, Publisher, Place of publication, Year, Language, Orig. language, ISBN, ISSN, NBN, Library symbol, Call number, Bib-ID.

In the Europeana portal, however, the searcher can continue the search by refining it by selecting one or more facets: Provider, Country, Type, Language, Date and Rights. So the search options are not as few as they look at first sight.

So to make precision search better through fielded search in a cross-domain search application a new kind of man-machine dialogue is needed. A dialogue that does not deny the heterogenous structure but that acknowledges it and helps the user to navigate through intelligent choices. For example, if only a fraction of the records have DDC classifications this must not make it impossible to use them in searching.
Free-text techniques

The other major approach to enhance precision is advanced free-text techniques such as truncations (left, right and center) and wild-cards for one or several characters, and there are also distance operators such as within a maximum number of words, in the same paragraph etc., and there is also the possibility to specify a certain order of terms. Here, too, combination of query-parts with boolean operators is common.

In cross-domain services the use of free-text to enhance precision is limited since the textual content have a very different role in for example a library database and a museum database. An abstract of a literary work and a description of a museum object, or an archeological site are in one sense similar and equivalent, but an abstract is limited in scope compared to the other descriptions. The same free text query will have a different meaning in the different domains.

Interoperability through mapping

Interoperability is defined in the IEEE Glossary as: the ability of two or more systems or components to exchange information and to use the information that has been exchanged. In a cross-domain service this can be interpreted to provide the same functionality for all the databases. An established way towards interoperability is mapping of various schemas to each other, for example mapping record fields onto Dublin Core.

Zeng and Chan (2006) give an overview of methodologies to achieve metadata interoperability. On the record level «efforts are intended to integrate the metadata records through the mapping of the elements according to the semantic meanings of these elements.» On the repository (collection) level efforts “focus on mapping value strings associated with particular elements (e.g. terms associated with subject or format elements). The results enable cross-collection searching.”

From the perspective of one individual domain the material from other domains can seem unstructured or incomplete; name forms can be different. All these differences are often considered to make the data “low quality”. From a cross-domain perspective this meaning of quality cannot be applied.

Within the European Digital Library [EDL] project http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/
a special study was carried out about metadata interoperability between archives, audio-visual archives, museums and libraries (Chambers, 2007). The European Library (TEL) developed what came to be called the TEL application profile. This was the foundation for the next phase: Europeana’s ESE, the Europeana Semantic Elements.It consists of the 15 original Dublin Core (DC) metadata elements plus a subset of the DCTerms and a set of thirteen elements which were created to meet Europeana’s needs. It is available at https://version1.europeana.eu/web/guest/technical-requirements/

Domain level metadata, Collection level and Object level metadata.

One thing that is lost in bringing together databases from different domains is the organization of information into specialized databases, collections. This organization enables a pre-selection of an information set. And the organization gives clues to relevance; for example, in the Medline database one can expect to find medical and health information.

One can also think of metadata on the domain level that can be valuable for searching, such as administrative metadata of various kinds that can give clues to content or context.

By recognizing the importance of collection level metadata, and maybe even domain level metadata and by including them in the database design further improvements in relevance can be achieved.

Semantic web technologies
Semantic web technologies can be of help to achieve relevance in searching – URI:s will, for example, bind together different name-forms for one and the same person. Rindoguren, Asutoriddo is perhaps difficult to recognize as a Japanese form for Линдгрен, Астрид or Astrid Lindgren; the URI ties them together: http://libris.kb.se/auth/71639. The National Library of Sweden has made the national union catalogue available as Linked Open Data.

In the Europeana v 1.0 project work has been done to develop a data model, EDM, that can bring semantic functionality to the portal. By carrying out a series of community meetings the model evolved by consensus among the cultural sectors: Libraries, AV archives, Museums, Archives. The latest version of the model is available at: http://version1.europeana.eu/web/europeana-project/technicaldocuments/

Precision can be improved by the semantic web, but harmonisation of the semantics is necessary. The EDM is a step towards this harmonization.

**Summing up**

The heterogeneity of structure which is inevitable in a cross-domain search service puts some constraints for the application of information retrieval techniques to improve relevance and precision.

Remedies can probably be found in techniques for semantic search, the semantic web, when it comes to improve relevance. User generated content in terms of commentaries and tags can also contribute. The latter can be expected to have positive effect on situational relevance.

For the improvement of precision fielded search cannot be applied in the traditional way because of the differences in record structure - a new kind of search dialogue is needed which is founded on the realization of this difference.

The use of free text methods to improve precision can be applied but this must be done with caution since the role of texts is different in the different domains. So there are still challenges to be met in order to achieve System Based Relevance.

The pursuit of relevance, both system based and user based, is of utmost importance for service designers since a service that cannot deliver relevant responses is soon abandoned.

**References and literature**


http://comminfo.rutgers.edu/~tefko/Saracevic%20relevance%20pt%20II%20JASIST%20%2707.pdf
[2011-04-15]
