

Sneak Preview? Instantly Know What To Expect In Faceted Browsing

Johannes Osterhoff, Jörg Waitelonis, Joscha Jäger, and Harald Sack
Hasso Plattner Institute for IT Systems Engineering
Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany

{johannes.osterhoff, joerg.waitelonis,
joscha.jaeger, harald.sack}@hpi.uni-potsdam.de

Abstract: This paper presents the Semantic Multimedia Explorer (SEMEX) a semantic search engine that supports exploratory video retrieval¹. SEMEX combines entities and content-based suggestions to support semantic search. The SEMEX interface provides facet filters, entity recommendations, a pagination and browsable search results; all of these elements are interconnected by instant “linking and brushing”. Thereby SEMEX enables instantaneous visual feedback on how a selected search facet will affect the displayed result set. This unique feature facilitates a more smooth and effortless search process compared to other systems that need to perform new search queries to achieve a similar result. By making the subsequent results available *at fingertip*, the SEMEX interface allows users to pursue related exploratory search strategies that lead to a genuinely responsive and quick browsing experience.

1 Introduction

With the introduction of *Google Instant*², popular web information retrieval has become quick. Immediately after entering a keyword into the search field, suggestions are shown in a drop down box and the result set is constantly updated during text input. Therewith *Google* provides instantaneous feedback on keyword search and resolves what once have been successively received web pages into a continuous flow of user input and system feedback. The video search of *Bing*³ for example, plays videos already in thumbnail preview and therewith transfers the users’ decision making process from the sites actually containing the video content to the page of the video result set itself.

While most keyword-based search systems are optimized to narrow down a huge data space to the most suitable results and present them in “ten blue links”, exploratory search

¹SEMEX was developed at the Hasso Plattner Institute for IT Systems Engineering in Potsdam, Germany in cooperation with Merz Akademie Stuttgart during the Mediaglobe project. Mediaglobe is part of the THESEUS research program, supported by the Federal Ministry of Economics and Technology on the basis of a decision by the German Bundestag.

²<http://www.google.com/instant/>

³<http://www.bing.com/videos/>

aims at finding results, which are not considered to be related at first glance. [WS11] Similar to the concept of the flaneur, an urban wanderer of the late 19th century, who leisurely walks through the more and more modernized urban landscape and thereby draws inspiration from, exploratory search assists the user in navigating the growing information spaces and lets her choose between alternatives, move along paths, and move back to discover sidetracks and to choose alternative ways. [DCW11] The growing amount of structured data on the web promises to enhance exploratory search by taking into account the actual meaning of the information and its semantics. An early system allowing its user to move through such information spaces and to quickly follow their “associative trails” is the famous *memex*. [Bus45] Envisioned over half a century ago by Vannevar Bush, it has been serving as model and distant goal for exploratory search. In concert with the ideas formulated with *memex*, the name SEMEX refers to Bush’s system.

Even though speed in terms of system feedback and an upfront but deep gaze into video result items have become available in popular web information retrieval, in the case of exploratory semantic search, instantness and an early preview into represented items on semantic level, is highly required but only slowly taking shape.

Exploratory semantic search is based on facets, enabling the user to better refine and broaden search queries and to provide content-based recommendations. [WKmcsS10] Extracting entities from vast data collections and to find meaningful representations on the interface is understood to be one of the most important challenges in search today. [BYBM11] In the field of semantic search, there is a fine line between usability and expressivity [WKW⁺10], user interaction overall needs to become more intuitive and *lightweight* and the success of exploratory semantic search heavily depends on accessible and user-friendly interfaces.

Taking these respects into consideration, this paper presents the semantic search engine SEMEX that facilitates its users to perform state-of-the-art exploratory video retrieval. The objective of the interface design was to provide its users with a quick feedback on selected facets that encourages exploratory search.

The paper is structured as follows: In Section 2 we describe related work and introduce the relevant technologies and paradigms. Section 3 deals with the realization of the user interface including design aspects. Then, Section 4 presents and discusses different show cases. Finally, Section 5 concludes the paper with a short discussion of achieved results and an outlook on future work.

2 Design Aspects and Related Systems

The Semantic Multimedia Explorer research project was initiated to develop technologies that reply to the growing amount of multimedia content on the World Wide Web. SEMEX combines state-of-the-art media analysis techniques, such as structural segmentation, intelligent character recognition, the ability to detect genres and faces in videos and semantic

technologies. For the current demonstrator videos from the *defa spectrum archive*⁴ have been analyzed; those contain documentaries on alternating subjects originating from the former German Democratic Republic. To make this content available to its users, the SEMEX interface makes use of an entity- and facet-based navigation to comprise and refine its search results.

In current interfaces, these elements often are represented by facet links with attached numbers. These numbers indicate how the result set will be refined when a link is selected. The academic video search engine “Yovisto”⁵ makes use of faceted browsing with a clear left-right-paradigm. [WS11] Users can expand their search queries by means of suggested DBpedia classes on the left. On the right, the current search can be filtered with facets derived from the results. The “CONTENTUS” interface, also made use of faceted browsing and provided a number behind each facet that indicated the change of the results after the filter would be applied. [WOS11] With “SemaPlover” it is possible to explore and visualize semantically heterogeneous distributed semantic data set in real time. Therefore the system also makes use of faceted search that are distinguished into locations, persons and tags. The changes take effect after a facet link is selected and no number or preview is given. [SSSS09] “Exhibit”⁶ is a toolkit to create faceted browsing interfaces. [HKM07] For example, it comes with various views, such as a time-line, and a map, which places the results geographically. Numbers associated to a facet link provide a hint, how the result set will be changed after selection. It is due to the nature of this toolkit that the result set can only be changed after explicit selection. The *Flamenco* search interface has the primary goal of allowing users to move through large information spaces like the painting collection of the San Francisco Fine Arts Museums or a Nobel Prize winners. *Flamenco* also uses numbers besides facet links and a tooltip to give a preview on consecutive facets. [?]

Stefaner et al published the elegant faceted search interface “Elastic Lists”. [SUS08] It supports searching from general to special terms, navigation by selection, and trial and error by search path backtracking. In most of their applications the “Elastic Lists” also provide a number behind their facet links to indicate how many results will be remaining after selection.

Even though all given examples provide a better way to refine searches than popular search engines do today, still they lack the immediacy of systems like *Google Instant* or the *Bing* video preview have to offer.

SEMEX makes use of “linking and brushing”, a principle commonly used to describe a connection between two or more views of the same data. In information visualization, this method is used to overcome the shortcomings of a standalone technique. [Kei02] “Linking and Brushing” was introduced by Becker and Cleveland [BC87] to brush scatterplot matrices. In that early case, brushing over a node on one graph highlighted related points on the other graphs.

For faceted refinements in a search process, this principle, also was used by Marian Dörk

⁴<http://www.defa-spektrum.de/>

⁵<http://testing.yovisto.com/>

⁶<http://www.simile-widgets.org/exhibit/>

et al with “VisGets” interconnecting various resources. [DCCW08] When brushing over elements such as tags, locations, time filters and thumbnails, the relation between these elements are highlighted. This example connects different elements in a meaningful way. Yet the pagination of the result set has not been enabled for brushing and “VisGets” does not allow the formulation of complex entity-based search queries.

The next section describes the SEMEX user interface in detail and points out its new features and their purpose.

3 The SEMEX User Interface

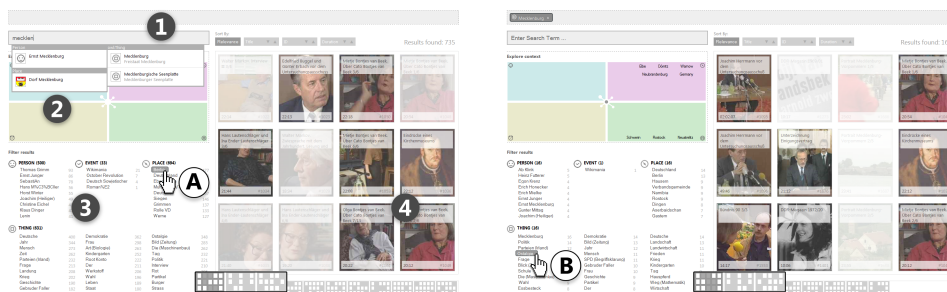


Figure 1: Brushing over different facets (A and B) highlights the linked thumbnails and changes their pagination on the right (4).

Considering Figure 1, the layout of the SEMEX user interface is arranged in a facet and search area (1), the *Facet Explorer* (2), facet filters (3) and the area for search results (4). The facet and search area spreads over the whole width of the layout. Below this area, on the left the *Facet Explorer* and the facet filters can be found. The largest area, used to display the search results, is located on the right.

3.1 Elements of the Interface

The *Context Explorer* consists of four colored squares, in which the SEMEX system suggests related entities divided into the four basic categories Persons, Places, Events and Things. **These suggestions are made possible by...**

Aligned in three columns below the *Context Explorer*, facet filters can be found; they are also grouped into four categories and are derived from the current search result. The facets used by SEMEX stem from the DBpedia.⁷

On the right hand's side the results are arranged in a tiled layout. Each result tile represents

⁷<http://www.dbpedia.org/>

a video and shows its title and duration, as well as a representative thumbnail.

3.2 Linking and Brushing Principle

Since selecting and de-selecting facets and suggestions is essential for the quick exploratory search of SEMEX, we wanted to provide an instant feedback on how selected suggestions or facets will affect the currently shown result set. When the user brushes over a facet link, as indicated in Figure 1 (A and B), linked thumbnails and tiles of the pagination that do not fit the criteria are grayed out (4). “Linking and brushing” also affects the pagination of the SEMEX and grays out results tiles that do not fit the selected criteria.

To implement “linking and brushing” of facets and result set...

3.3 Demonstration Screencast

Because of copyright protection on its content, a public live demo⁸ of SEMEX can only be provided at the workshop. Anyway, a screencast demonstrates the user interface in all details (c.f. <http://www.bit.ly/semexdemo>). The following section describes the capabilities of the SEMEX user interface by means of selected search scenarios.

4 Search Scenarios and Demonstration

4.1 First Queries and Disambiguation

This example starts with typing “Mecklenburg”⁹ into the search mask of SEMEX. While typing, a disambiguation matrix is displayed, that suggests what the term entered stands for. In our case Mecklenburg might refer to a person called Ernst Mecklenburg, a place, such as the village Mecklenburg, or other entities, for example the German federal state. The suggested entities are grouped in different blocks and can be recognized by distinct icons. When selecting Freistaat Mecklenburg, the German federal state, the search term is confirmed and from now on is listed as an entity in the facet area on top of the page. The items listed in this area comprise active entities and search terms that determine the shown search results.

⁸<http://mediaglobe.yovisto.com:8080/mggui/>

⁹The examples given are due to the content of the *defa spektrum archive*.

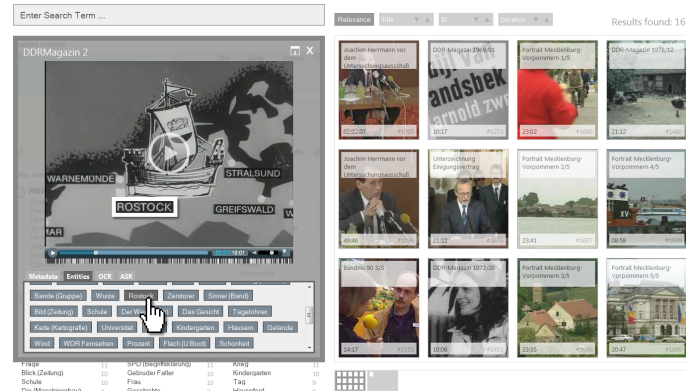


Figure 2: Detailed view of the result video: selecting an entity directs to its occurrence in the video.

4.2 Context Explorer and Facet Filters

On the left hand's side the *Context Explorer* and the facet filters can be found. The *Context Explorer* consists of four colored squares, in which the SEMEX system suggests related entities divided into the four basic categories. For this search, SEMEX for example suggests Schwerin, which is the capital of Mecklenburg, or Rostock, its largest city. When brushing over Rostock, about half the results remain visible; when brushing over Schwerin, only one result remains.

Aligned below the *Context Explorer* the facet filters are also grouped into the four categories. These facets are derived from the current search result. In the Person category for example, "Erich Honecker" can be found, the political leader of the German Democratic Republic. He is accompanied by "Egon Krenz", Honecker's successor. In the category for Things "Ostalgie" can be found, a German term referring to nostalgia for aspects of life in East Germany. When brushing over the two politicians, four results remain visible, nine when brushing over "Ostalgie".

By making the facets available for brushing and linking them to the representative thumbnails, SEMEX provides valuable feedback on how the result set will be changed according to the next input provided by the user. Eventually, a click on a facet, places it in the facet area on top of the page and the results are changed accordingly.

4.3 Video Details

By selecting an item within the search results, the left column is overlaid with information about the selected video (see Figure 2). There also the video player is located. In a tabbed layout below the player, entities detected by Automated Speech Recognition and Video OCR are listed. When an entity is selected, the video jumps to the occurrence of the entity

in the current video.

5 Conclusion and Future Work

The search with SEMEX cannot be achieved with current search engines. Semantic entities enable video search beyond the capabilities of keywords and the quick refinement of the search query and the preview made possible by “linking and brushing” facilitates the user to perform searches that are not only exploratory but also *lightweight*. The linking of facets and elements like thumbnails and pagination is a real benefit for exploratory search – leaving behind the paradigm of consecutively loaded pages making way for quick preview and deep gaze into the content of videos.

In future work we plan to investigate, how further elements of the interface can be put into meaningful interconnections to each other. For example, brushing could also be applied to segments of the videos, the entities of the detail view could be connected back to the search results and the pagination could make use of different display methods.

6 Acknowledgments

The SEMEX was developed during a joint project of Hasso Plattner Institute for IT Systems Engineering and Merz Akademie Stuttgart. The authors would like to thank to Mario Doulis, Jörg Frohnmayer and Stephan Schröter for their valuable feedback and contribution to the project.

6.1 Weitere mögliche Titel

Sneak Preview... Explorative Semantic Search with Faceted Suggestions at Your Fingertip
Sneak Preview... Gaining Insights into Results before Selecting them
Sneak Preview...
Instantly Know What To Expect In faceted Exploratory Search

References

- [BC87] Richard A. Becker and William S. Cleveland. Brushing scatterplots. *Technometrics*, 29:127–142, May 1987.
- [Bus45] Vannevar Bush. As We May Think. *Atlantic Monthly*, 176(1):641–649, March 1945.
- [BYBM11] Ricardo Baeza-Yates, Andrei Broder, and Yoelle Maarek. The New Frontier of Web Search Technology: Seven Challenges. In Stefano Ceri and Marco Brambilla, editors,

Search Computing, volume 6585 of *Lecture Notes in Computer Science*, pages 3–9. Springer Berlin / Heidelberg, 2011.

- [DCCW08] Marian Dörk, M. Sheelagh T. Carpendale, Christopher Collins, and Carey Williamson. VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1205–1212, 2008.
- [DCW11] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. The Information Flaneur: A Fresh Look at Information Seeking. CHI 2011 Proceedings, 2011.
- [HKM07] David Huynh, David Karger, and Rob Miller. Exhibit: Lightweight Structured Data Publishing. In Carey L. Williamson, Mary Ellen Zurko, Peter F. Patel-Schneider, and Prashant J. Shenoy, editors, *Proceedings of the 16th International World Wide Web Conference*, pages 737–746, Banff, Alberta, May 2007. ACM Press.
- [Kei02] Daniel A. Keim. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8:1–8, January 2002.
- [SSSS09] Simon Schenk, Carsten Saathoff, Steffen Staab, and Ansgar Scherp. SemaPlorer – Interactive Semantic Exploration of Data and Media based on a Federated Cloud Infrastructure. *Journal of Web Semantics*, 7(4):298–304, 2009.
- [SUS08] Moritz Stefaner, Thomas Urban, and Marc Seefeldler. Elastic Lists for Facet Browsing and Resource Analysis in the Enterprise. In *Proceedings of the 2008 19th International Conference on Database and Expert Systems Application*, pages 397–401, Washington, DC, USA, 2008. IEEE Computer Society.
- [WKmcsS10] Max L. Wilson, Bill Kules, m. c. schraefel, and Ben Shneiderman. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science*, 2(1):1–97, 2010.
- [WKW⁺10] Jörg Waitelonis, Magnus Knuth, Lina Wolf, Johannes Hercher, and Harald Sack. The Path is the Destination – Enabling a New Search Paradigm with Linked Data. In *Proceedings of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly, December 16-17, 2010, Ghent, Belgium, CEUR Workshop Proceedings*, volume 700, 2010.
- [WOS11] Jörg Waitelonis, Johannes Osterhoff, and Harald Sack. More than the Sum of its Parts: CONTENTUS – A Semantic Multimodal Search User Interface. volume 694 of *Proceedings of the Workshop on Visual Interfaces to the Social and Semantic Web*, January 2011.
- [WS11] Jörg Waitelonis and Harald Sack. Towards exploratory video search using linked data. *Multimedia Tools and Applications*, pages 1–28, 2011. 10.1007/s11042-011-0733-1.