Jul 9th, 12:00 AM

Mobile discovery for libraries and museums

Wolfgang Stille
Darmstadt University and State Library, Germany

MOBILE DISCOVERY FOR LIBRARIES AND MUSEUMS

Wolfgang Stille
University and State Library Darmstadt
TU Darmstadt, Germany
stille@ulb.tu-darmstadt.de

Abstract. In 2014, over one billion smartphones have been sold worldwide, around 3 trillion search queries were entered solely on Google, a growing part from mobile devices. With rapidly growing digitization, enormous changes accompany almost all areas of our lives and define fundamentally new ways we deal with information, education, and culture. At the same time, cultural institutions increasingly assume social responsibilities that go beyond the mere mediation of knowledge. For this purpose, infrastructures are necessary that allow to archive digital content independent from their media type and domain and make it available to the public; on the other hand, it is necessary to semantically link this content and present it to a more and more demanding audience that is already used to interactive and participatory forms of knowledge dissemination and information research. The cloud media projects in Hesse created a platform to give access to heterogeneous content types and domains and link them semantically. Two strategic partners, the Städel Museum in Frankfurt and the Darmstadt University and State Library, provide use cases for 2 applications based on the platform: a discovery service for literature, and a virtual museum application for objets d'art, both implemented with responsive design and touch screen support to enable mobile usage. The first provides a search and discovery in the library's resources using a topic pie to support an efficient and intuitive drill-down of extensive search results. The second allows an exploration of the Städel's collection on mobile devices; users are able to find similar artworks considering different features such as artist, theme, motif, era, atmosphere, etc. Both take advantage of touch-sensitive displays and allow an intuitive and ludic handling. They break fresh ground by leaving the classic search and find paradigm towards digital strolling: just like a physical visit to a library or museum, strolling around, discovering things, and getting inspired.

Keywords: Discovery Services, Digital Strolling, Explorative Search, Mobile Devices, Responsive Design, Cultural Heritage.

In 2012, University and State Library at TU Darmstadt started a multi-year project with several partners aiming at mobile discovery in order to extend the physical space of cultural heritage institutions such as libraries and museums. One of the project's foci is the exploration of an object space leaving the classic search and find paradigm we are used to when entering the digital world and moving towards more explorative and intuitive scenarios. Explicitly, the project did not aim at emulating or transferring the onsite experience of an exhibition or library space onto a tablet computer. Instead the goal was to exploit the benefits of the digital age, such as unbounded presentation space, or independency of time and space, and combine them with the digital support of behaviour that visitors usually display in cultural heritage institutions: being curious, strolling around, discovering things, and getting inspired. A major benefit hereby is that we could reveal objects that are hidden in magazines and usually not accessible to the public; another is that we could draw connections between all of the objects in the pool despite of their physical absence. The second important aspect is the support of mobile devices as they play a steadily increasing role in the zoo of devices people utilize for making use of cultural heritage institutions' online services. Thus, a major requirement was the responsive design of the presentation platform's applications and the support of touch sensitive displays in order to allow an intuitive handling on all kinds of devices. Finally, the enrichment of objects by means of surrounding and semantically related information such as artist biographies, audio guides, or explainity videos was a major goal of the project. The storage of objects of different use cases on the same platform also allows semantic linkage between those; so we could recommend literature from the library in the museum application and vice versa.

The project provides a scalable and multi-client capable cloud-based architecture that might be deployed in a virtual server environment. That allows a very flexible usage of the platform: as an autonomous and standalone operation in large cultural heritage institutions on the one hand, or

as a software as a service solution (SaaS) hosted and maintained by a professional IT provider on the other hand. This might be interesting in particular for smaller and medium size institutions lacking own IT facilities.

Consortium

The project combines the expertise from four different areas - science, engineering, cultural heritage, and business. Therefore, the consortium consists of several partners from science, cultural heritage institutions, and industry. Scientific partner in the first place is Hochschule Darmstadt, University of Applied Sciences, which is involved with two research groups, *Software Engineering* and *Development of Media Systems*. Main focus lies on the application of semantic technologies, the development of a prototype platform in close collaboration with the application and industry partners, and the implementation of modern user interfaces. Also the department of electronic information services of University and State Library Darmstadt contributed scientific input and its realisation to both, the library as well as the museum platform. Application partners are the University and State Library of TU Darmstadt, and the Städel Museum in Frankfurt. They constituted the two use cases - modern library services and digital exploration of art. Apart from all the requirements, they contributed all the data and subject-specific expertise, e.g. how curation in the digital world works, or what kinds of metadata are useful for added value services. Partners from industry are Software AG, the second largest software vendor in Germany, media transfer AG, an expert in the field of user interfaces, nterra GmbH, an IT consulting company, and House of IT, a center focusing on the promotion of innovative information and communication technologies. They will operate the platform as a software as a service solution in the future.

Use cases

Two use cases were identified in the project proposal and cultivated during the whole lifetime of the project: library and museum. While the purpose of the museum application is to establish and support the *self-conducted process of learning and enjoying the jump into the rich and complex world of human nature seen through art* (Eschenfelder et al., 2013) in the digital world with a thoroughly educational claim, the library application supports the navigation through large literature collections along different dimensions in order to find relevant objects efficiently. Both applications support an explorative search, i.e. finding without searching which we called *digital strolling*. This is achieved by semantic linking of related objects on the one hand, and the implementation of user interfaces that allow intuitive navigation through the collections on the other hand. We will illuminate the process for both, the library and the museum application in detail below.

Data

One of the major challenges in the project was data: how to choose the data model for the sake of completeness and consistency, but expandability in the future? What data to aggregate from which existing sources, and – even more challenging – how to collect data that is not yet present but strongly required for the use of semantic technologies, such as for example sentiments connoting a certain objet d'art. Similar to other projects like (Bermès, 2014), we decided to establish our own data model. The result was a quite generic and extendible data model, that is loaded by an ETL (extraction, transformation, loading) process from different existing databases: the library catalogue, the museum's adlib database, and data from Gemeinsame Normdatei (GND), a common authority file of named entities such as persons, subjects, places, and organizations that is coordinated by the German National Library. The extraction step acquires data from relevant sources of sufficient quality. The information provided is saved in structured and machine-readable formats such as MARC or RDF. The transformation step does not only comprise format conversion, but in addition to this semantic enrichment e.g. by matching corresponding entities from different data sources via heuristic methods (cf. Heuss et al., 2013). In the last step, data is loaded into a data store that provides sufficient query functionality and performance such as an index. In the project SOLR/Lucene is used for that purpose.

While metadata of library objects was present and reviewed for the most part, the situation was completely different in the museum case: except for essential data such as artist and title, much information had to be collected during the first project phase. For this reason, a team of art history experts and students tagged work by work manually. In the second and currently ongoing project phase, we think about (semi-)automatic text mining techniques, or crowd sourcing by the users – but only as an addendum to intellectually generated metadata.

Semantic linking and user interfaces

The strength of the two developed applications is the combination of semantic techniques combined with appropriate user interfaces and guidance logics. We will illustrate this in detail in the following. For the library application, we combine an advanced faceting algorithm, that overcomes the common drawbacks of faceting, with a visualisation technique that optimally suits mobile devices and allows an intuitive navigation through hierarchical facets in order to drill down the search result set even on small displays. For this purpose, we make use of a graphical element as an interactor for the topical faceting, referring to a data visualisation technique recommended by (Collins, Carpendale, & Penn, 2009). We adapted the pie menu to visually represent two levels of a hierarchical topic taxonomy and called it a *topic pie*. It allows users to refine their search results by means of two layers of hierarchical information given by a taxonomy of topics. The outer ring of the element resembles topics which themselves again contain a topically similar document collection each. The inner ring of the pie menu contains terms, which label an aggregation of several topics from the outer ring into more general topics, such that the hierarchical taxonomy is preserved. Figure 1 shows the generation of the topic pie from the taxonomy. Isochromatic elements from the taxonomy on the left hand side match the colours in the pie menu on the right hand side.

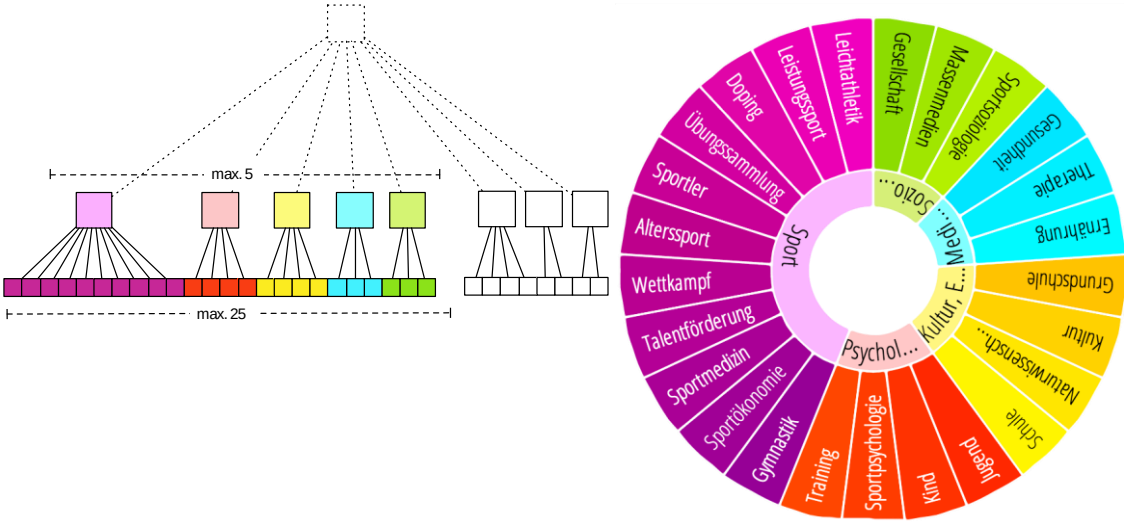

Figure 2: Hierarchical topic pie built from a taxonomy tree (topics in German)

We use topic pies to implement topical faceting of search result sets, that is, to explore and drill down a set of search results from a given query. The benefits of showing two hierarchical levels at the same time is twofold: on one hand, it enables the user to navigate towards fine-grained topics, whilst a more coarse-grained level shows the coherence of topics on an upper level. On the other hand, it allows disambiguation of ambiguous search terms. A click on a specific pie segment selects the corresponding topical facet, and thus results in a smaller search result set. A new topic pie is generated according to the new result set. The procedure allows users to refine their search by iteratively drilling down search results. Figure 2 illustrates the drill down of the faceted navigation for the initial search request ”Java” (3257 results, topic pie (a) on the left hand side), which mainly displays Computer Science topics (shown in blue), but also others outside this subject area, for example related to the island in Indonesia (the yellow sector). In order to narrow down these results, the user may choose a specific topic from the outer ring, (e.g., Web Services) or the whole sector (e.g. Computer Science) by selection of the inner ring. After choosing the domain Computer Science the search result set is reduced (1536 results, topic pie (b) in the middle). The pie adapts accordingly by presenting more topics from this domain. Finally selecting the topic ”Web Services” reduces the result set again (23 results, topic pie (c) on the right hand side).

Technically, the topics used in our approach are taken from existing metadata of the object collection, i.e. subject terms that are mapped and aggregated on a taxonomy derived from GND. Alternatively, metadata might be generated from full text using text mining technologies, for example topic models or named entity recognition systems. For details of the whole algorithmic

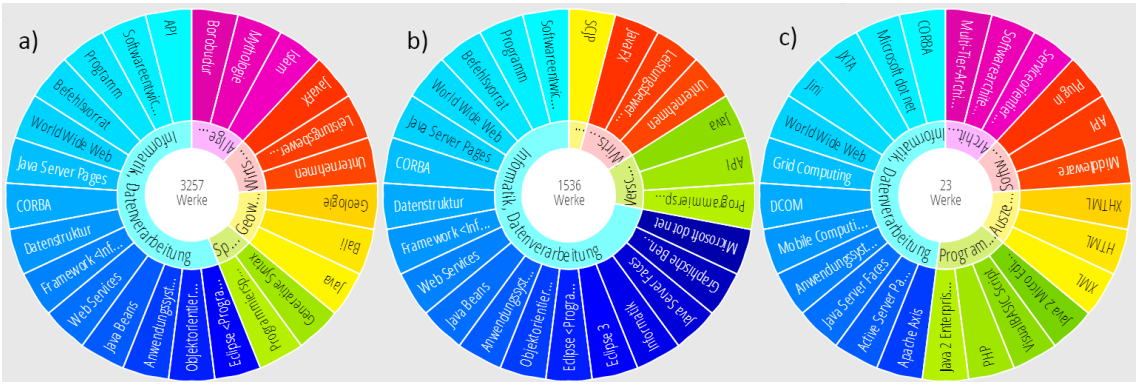

Figure 2: Sequence o f topic pies during an iterative drill-down process: (a) pie from search term ”Java”, (b) after selection of ”Informatik, Datenverarbeitung” in the inner blue ring, and (c) after selection of “Web Services” on the outer wheel.

process, see our work in (Deuschel, Greppmeier, Humm, & Stille, 2014).

We also did a user evaluation of our mobile application using the topic pie by qualitative and quantitative means with librarians and university students. As a result, the topic pie shows major advantages of the approach compared to state-of-the-art faceted search techniques in discovery services. The selectivity of generated topic pies was checked by a test similar to (Chang, Gerrish, Wang, & Blei, 2009) in order to ensure that the recommended facets are homogeneous and meaningful to the given search term. For further details of the evaluation confer to (Deuschel, Greppmeier, et al., 2014).

Alternatively, the topic pie might be used in order to show a flat, but multi-dimensional faceting. In a prototype, we show a topical facet and a location facet. The choice of one of these facets again affects the result set and therefore the generation of the new topic pie. Figure 3 illustrates the interaction between facets and search result set. This variant of the topic pie does not allow a successive drill-down of search results, but it allows the choice of more than one facet and more than one entry per facet at once instead.

Another option is the exploration of the library catalogue without a concrete search term, but just by using the topic pie. Technically, this resembles a search for ‘*’ and faceting via the hierarchical or flat topic pies. From a usability point of view, it is similar to a walk to the shelf of a specific area of subject in the physical library. As a matter of course, the application contains all elements of state-of-the-art discovery services, too.

Digital strolling is also featured in the museum application, but in a much more prominent manner. Here, we use many types of relations in order to present similar or related content to the user, e.g. with respect to era, impression, painting technique, material, connotations, pictorial elements, atmosphere, etc. Visualized links show the relation between related objects,

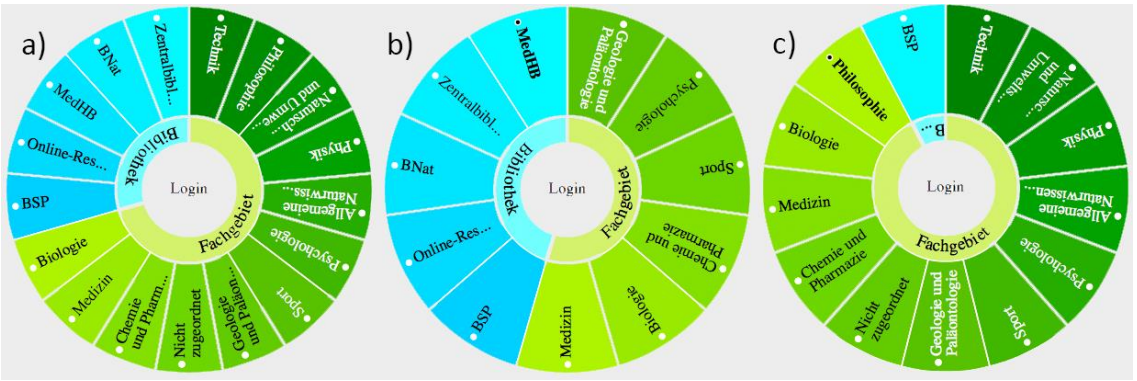

Figure 3: 2-dimensional topic pies showing location (blue) and subject (green) facets: (a) after search input ”Mikrotubuli”, (b) after selection the medical library on the location facet, and (c) after selection of philosophy on the subject facet.

and the platform usually recommends related objects with respect to a variety of different relations. Along these link paths, the users are able to move in different directions and explore the whole collection in this manner. The paths are saved and displayed, so entering any arbitrary position on the path is possible at any time. This way of digital strolling is complemented by classic extended search functionality, a so-called *chocolate box*, where users

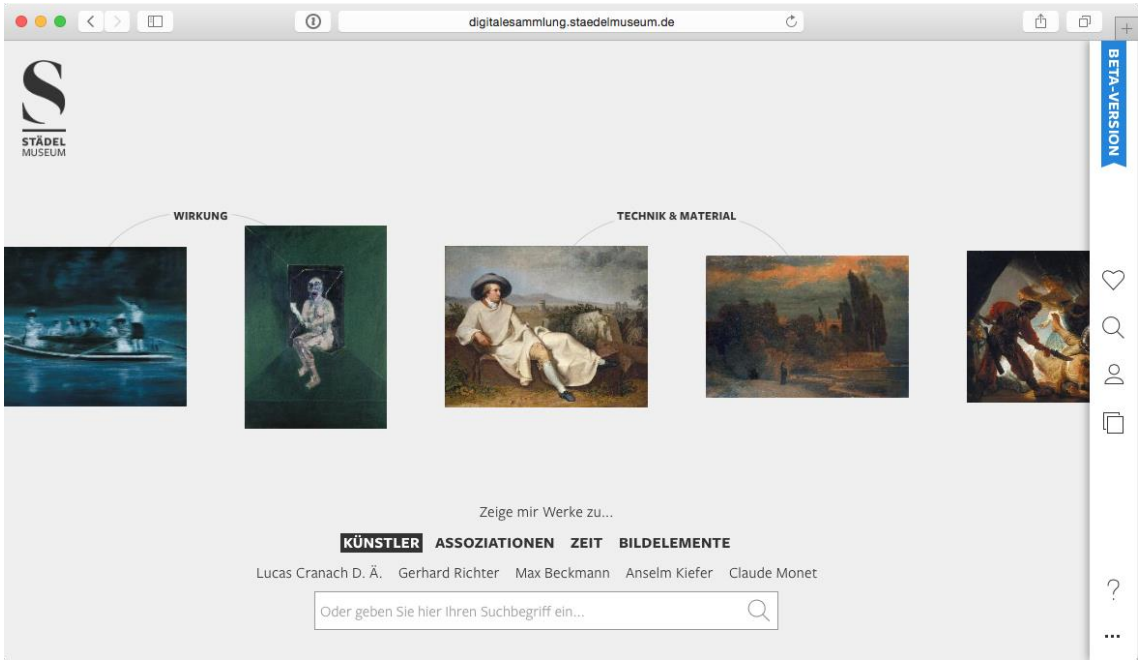

Figure 4: Entry screen of the 'Städel Digitale Sammlung' showing a choice of paintings and typed relations between them. Alternatively, a search may be entered.

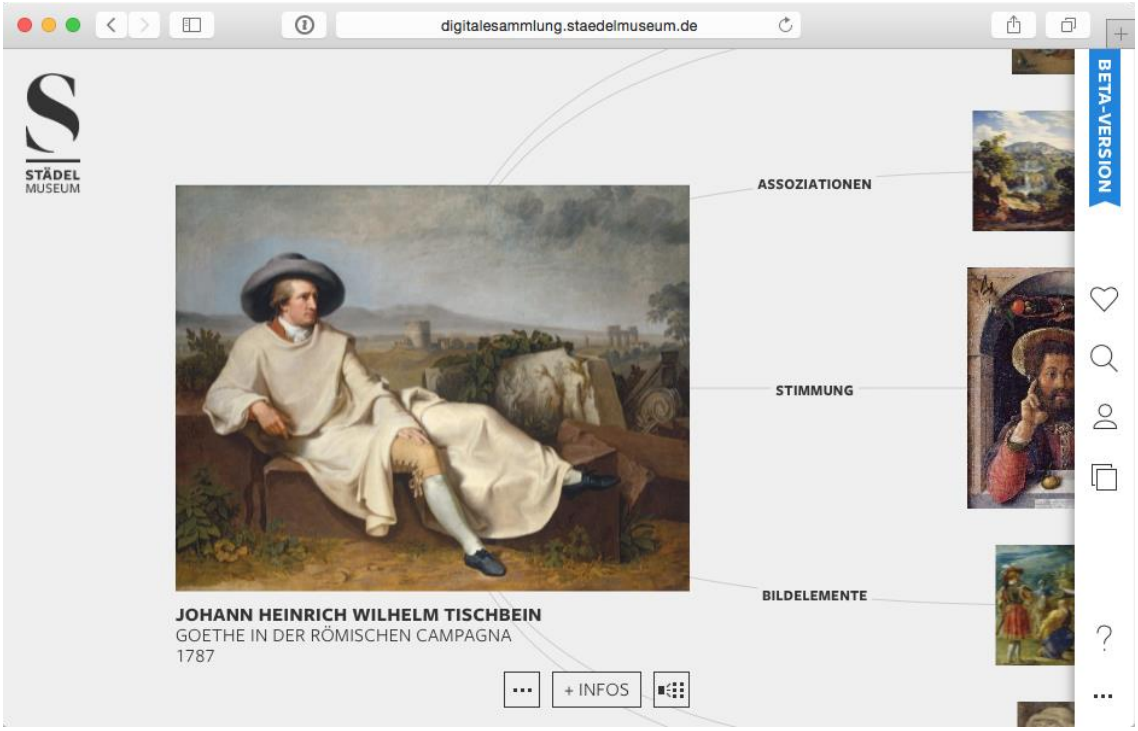

Figure 5: From the latest viewed object, paths of different types lead to recommended related objects and allow serendipity. Supplemental information and a full screen view might be picked by pointing or clicking on the object.

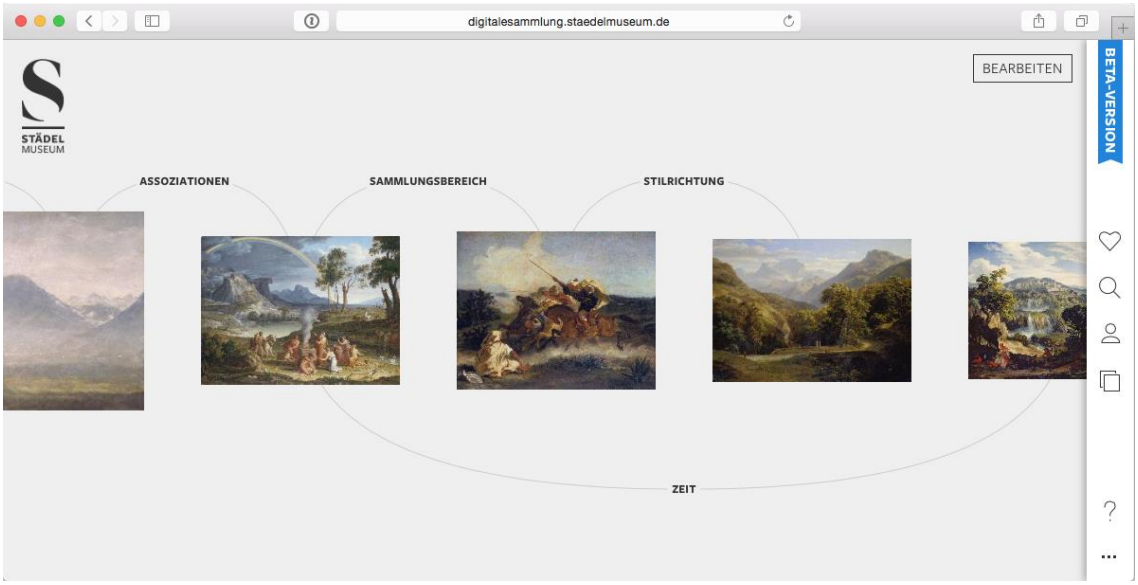

Figure 6: Reviewing the path and its relations and editing it is possible at any time.

are able to put in and share their favourite objects, and curated albums. For any object, supplemental information and material are offered such as picture description, history, and different audio and video content. Figures 4 to 6 show screenshots of the current application during its beta phase illustrating the features mentioned above.

Conclusion and Perspective

In this survey paper, we gave an overview of an ongoing project that has he objective to support presentation, exploration, and curation of digital object spaces on mobile devices with high demands on usability and aesthetics. For the library application, we presented a new topical faceting approach using a graphical menu element called topic pie. The user evaluation has shown major advantages of faceted search using the topic pie. For this reason, we will embed it in our productive mobile discovery service soon. For the museum application, we illustrated a system that is already in beta phase. It is planned to offer the framework as a software as a service solution, that enables also smaller museums und cultural heritage institutions to present their collections in a modern and attractive manner. The consortium displayed its strength by a well balanced combination of partners from science and industry, caring for continuation and utilization of the project's results. Two early adopters as 'customers' delivering distinct use cases complement the picture. Further use cases and adaptations of user interfaces will follow.

Ongoing research is currently done in the areas of user understanding, recommender systems, and digital storytelling (Deuschel, Heuss, & Broomfield, 2014). We are aware that personalized services are a consequent way of further development to accomplish state of the art applications for the illustrated use cases. Currently, we are collecting session data from both applications with the goal to build user models in order to make personalized recommendations in the future. We also identified digital storytelling as an important variant of knowledge mediation, in particular to get young people interested in cultural heritage. Integration of social media is an important topic, which we already support in one direction: sharing content with other users via social media platforms. The other direction, integration of social media content, and crowd sourcing and tagging, e.g. in order to collect index terms from everyday language to enrich the vocabulary and objects' metadata, is still pending.

One important lesson learned during the project is that the homogeneity and quality of data is one of the most important things. Therefore, complementing activities increasing the coverage and quality of data are essential for added value services like this. As intellectual assessment of metadata is expensive and not always affordable with respect to the steadily growing amount of data, (semi-)automatic ways are necessary to achieve this. In general, text mining and natural language processing techniques might be used for that purpose, either standalone and supplemental to intellectually generated index terms, or as part of a semi-automatic process, where qualified staff decides in the end which of the proposed automatically extracted terms to

choose, similar to (Malits & Schäuble, 2014). A first step in the project has been undertaken by evaluating appropriate systems (Heuss, Humm, Henninger, & Rippl, 2014). Projects matching different vocabularies and classification schemes such as (Balakrishnan, 2013) must be also integrated into this process.

Ackowledgments. This work has been funded by the *Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz* (LOEWE) of the state Hesse, Germany, under grants HA 320/12-10, 320/12-11, 321/12-11, and 417/07. Thanks to all project partners for their contributions and the successful collaboration that made this project possible.

References

Balakrishnan, U. (2013). Das VZG-Projekt coli-conc -- Brückenbildung zwischen DDC und RVK. In *RVK-Anwendertreffen 2013*. Retrieved from http://rvk.uni-regensburg.de/images/stories/Conf2013/bal

Bermès, E. (2014). Following the user's flow in the Digital Pompidou. *Zeitschrift Für Bibliothekswesen Und Bibliographie*, *61*(2), 68–75.

Chang, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 288–296). http://doi.org/10.1.1.100.1089

Collins, C., Carpendale, S., & Penn, G. (2009). DocuBurst: Visualizing document content using language structure. *Computer Graphics Forum*, *28*(3), 1039–1046.

Deuschel, T., Greppmeier, C., Humm, B. G., & Stille, W. (2014). Semantically faceted navigation with topic pies. In *Proceedings of the 10th International Conference on Semantic Systems - SEM '14* (pp. 132–139). New York, New York, USA: ACM Press. http://doi.org/10.1145/2660517.2660522

Deuschel, T., Heuss, T., & Broomfield, C. (2014). The Design Scope of Adaptive Storytelling in Virtual Museums. In R. Klein & P. Santos (Eds.), *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association. http://doi.org/10.2312/gch.20141308

Eschenfelder, C., Gresch, K., Fröhlich, T., Humm, B., Greiner, T., Eierdanz, P., & Blumenberg, F. (2013). The other way round: from semantic search to collaborative curation. In *Proceedings of the Nordic Digital Excellence in Museums Conference {(NODEM 2013)}, Stockholm, Sweden, December 2013.* Retrieved from http://sched.co/HYjr6V

Heuss, T., Humm, B., Deuschel, T., Fröhlich, T., Herth, T., & Mitesser, O. (2013). Semantically Guided, Situation-Aware Literature Research. In *Workshop on User Interaction built on Library Linked Data (UILLD 2013), 79th World Library and Information Conference, Singapore*.

Heuss, T., Humm, B., Henninger, C., & Rippl, T. (2014). A Comparison of NER Tools W.R.T. A Domain-specific Vocabulary. In *Proceedings of the 10th International Conference on Semantic Systems* (pp. 100–107). New York, NY, USA: ACM. http://doi.org/10.1145/2660517.2660520

Malits, A., & Schäuble, P. (2014). Der Digitale Assistent: Halbautomatisches Verfahren der Sacherschließung in der Zentralbibliothek Zürich. *ABI Technik*, *34*(3-4), 132–143. Retrieved from http://www.degruyter.com/view/j/abitech.2014.34.issue-3-4/abitech-2014-0024/abitech-2014-0024.xml