

5. RELATED WORK

Multimodal search can be used in two senses; (i), in the sense of multimodal result output based on unimodal query input, and (ii), in the sense of multimodal result output *and* multimodal query input. We follow the second definition, i.e., require the query input interface to allow for multimodality.

An interesting multimodal search engine was developed in the scope of the PHAROS project [4]. With the initial query being keyword-based, content-based or a combination of these, the search engine allows for refinement in form of facets, like location, that can be considered modalities. I-SEARCH develops this concept one step further by supporting multimodality from the beginning. In [8], Rahn Frederick discusses the importance of multimodality in search-driven on-device portals, i.e., handset-resident mobile applications, often preloaded, that enhance the discovery and consumption of endorsed mobile content, services, and applications. Consumers can navigate on-device portals by searching with text, voice, and camera images. Rahn Frederick’s article is relevant, as it is specifically focused on mobile devices, albeit the scope of I-SEARCH is broader in the sense of also covering desktop devices. In a W3C Note [11], Larson *et al.* describe a multimodal interaction framework, and identify the major components for multimodal systems. The multimodal interaction framework is not an architecture *per se*, but rather a level of abstraction above an architecture and identifies the markup languages used to describe information required by components and for data flows among components. With Mudra [10], Hoste *et al.* present a unified multimodal interaction framework supporting the integrated processing of low level data streams as well as high level semantic inferences. Their architecture is designed to support a growing set of input modalities as well as to enable the integration of existing or novel multimodal fusion engines. Input fusion engines combine and interpret data from multiple input modalities in a parallel or sequential way. I-SEARCH is a search engine that captures modalities sequentially, however, processes them in parallel.

6. FUTURE WORK AND CONCLUSION

The efforts in the coming months will focus on integrating the different components. Interesting challenges lie ahead with the presentation of results and result refinements. In order to test the search engine, a set of use cases has been compiled that covers a broad range of modalities, and combinations of such. We will evaluate those use cases and test the results in user studies involving customers of the industry partners in the project.

In this paper, we have introduced and motivated the I-SEARCH project and have shown the involved components from the different project partners. We have then presented first results, provided a system demonstration, and positioned our project in relation to related work in the field. The coming months will be fully dedicated to the integration efforts of the partners’ components and we are optimistic to successfully evaluate the set of use cases in a future paper.

7. ACKNOWLEDGMENTS

This work was partially supported by the European Commission under Grant No. 248296 FP7 I-SEARCH project.

8. ADDITIONAL AUTHORS

Additional authors: Jonas Etzold, Paul Grimm (Hochschule Fulda, email: {jonas.etzold, paul.grimm}@hs-fulda.de), Athanasios Mademlis, Sotiris Malassiotis, Petros Daras, Apostolos Axenopoulos, Dimitrios Tzovaras (CERTH/ITI, email: {mademlis, malasiot, daras, axenop, tzovaras}@iti.gr)

9. REFERENCES

- [1] Apple iPhone 4S – Ask Siri to help you get things done. Avail. at <http://www.apple.com/iphone/features/siri.html>.
- [2] Google Voice Actions for Android, 2011. Avail. at <http://www.google.com/mobile/voice-actions/>.
- [3] Google Voice Search – Inside Google Search, 2011. Avail. at <http://www.google.com/insidesearch/voicesearch.html>.
- [4] A. Bozzon, M. Brambilla, Fraternali, et al. PHAROS: an audiovisual search platform. In *Proceedings of the 32nd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, SIGIR ’09, pages 841–841, New York, NY, USA, 2009. ACM.
- [5] P. Daras, A. Axenopoulos, V. Darlagiannis, et al. Introducing a Unified Framework for Content Object Description. *Int. Journal of Multimedia Intelligence and Security (IJMIS)*. Accepted for publication. Avail. at <http://www.lsi.upc.edu/~tsteiner/papers/2010/rucod-specification-ijmis2010.pdf>, 2010.
- [6] A. Emtage, B. Heelan, and J. P. Deutsch. Archie, 1990. Avail. at http://archie.icm.edu.pl/archie-adv_eng.html.
- [7] J. Etzold, A. Brousseau, P. Grimm, and T. Steiner. Context-aware Querying for Multimodal Search Engines. In *18th Int. Conf. on MultiMedia Modeling (MMM 2012)*, Klagenfurt, Austria, January 4-6, 2012. <http://www.lsi.upc.edu/~tsteiner/papers/2012/context-aware-querying-mmm2012.pdf>.
- [8] G. R. Frederick. Just Say “Britney Spears”: Multi-Modal Search and On-Device Portals, Mar. 2009. Avail. at <http://java.sun.com/developer/technicalArticles/javame/odp/multimodal-odp/>.
- [9] A. Gruenstein, I. McGraw, and I. Badr. The WAMI Toolkit for Developing, Deploying, and Evaluating Web-accessible Multimodal Interfaces. In *ICMI*, pages 141–148, 2008.
- [10] L. Hoste, B. Dumas, and B. Signer. Mudra: A Unified Multimodal Interaction Framework. 2011. Avail. at <http://wise.vub.ac.be/sites/default/files/publications/ICMI2011.pdf>.
- [11] D. R. James A. Larson, T.V. Raman. W3C Multimodal Interaction Framework. Technical report, May 2003. Avail. at <http://www.w3.org/TR/mmi-framework/>.
- [12] P. Letessier, O. Buisson, and A. Joly. Consistent visual words mining with adaptive sampling. In *ICMR*, Trento, Italy, 2011.
- [13] G. Little. Smiley Slider. Avail. at <http://glittle.org/smiley-slider/>.