One Size Does Not Fit All – Multimodal Search on Mobile and Desktop Devices with the I-SEARCH Search Engine

Thomas Steiner
Google Germany GmbH
tomac@google.com

Lorenzo Sutton
Accademia Naz. di S. Cecilia
l.sutton@santacecilia.it

Sabine Spiller
EasternGraphics GmbH
spiller@easterngraphics.com

Marilena Lazzaro,
Francesco Nucci, and
Vincenzo Croce
Engineering
{first.last}@eng.it

Alberto Massari, and
Antonio Camurri
University of Genova
alby@infomus.dist.unige.it,
antonio.camurri@unige.it

Anne Verroust-Blondet,
and Laurent Joyeux
INRIA Rocquencourt
{anne.verroust,
laurent.joyeux}@inria.fr

ABSTRACT

In this paper, we report on work around the I-SEARCH EU (FP7 ICT STREP) project whose objective is the development of a multimodal search engine targeted at mobile and desktop devices. Each of these device classes has its specific hardware capabilities and set of supported features. In order to provide a common multimodal search experience across device classes, one size does not fit all. We highlight ways to achieve the same functionality agnostic of the device class being used for the search, and present concrete use cases.

Categories and Subject Descriptors

H.3.4 [**Information Systems**]: Information Storage and Retrieval—*World Wide Web*

General Terms

Experimentation

Keywords

Multimodality, Rich Unified Content Description, IR

1. INTRODUCTION

Even in 2012, search is a mainly text-driven operation. Albeit recent developments in the mobile and desktop worlds have introduced voice search as an *additional* input modality, a *truly multimodal* search experience with multimodal in- and output is still missing. In the scope of I-SEARCH, industrial and academic partners are working together to investigate ways to provide such a multimodal search experience across device classes. We provide an overview of the project in its entirety in [11]. An important step towards multimodal search in the scope of I-SEARCH was the creation of a unified annotation format named *Rich Unified Content Description (RUCoD)*, which was detailed in [4].

In this paper, we focus on taken actions and future plans to deal with device constraints to support the in- and output modalities *audio*, *video*, *rhythm*, *image*, *3D object*, *sketch*, *emotion*, *geolocation*, and *text*. The I-SEARCH project is in its second year now, and some basic functionality is in place. We maintain a demonstration server[1], and have also recorded a screencast[2] showing features of the search engine.

2. USER-FACING PROJECT GOALS

I-SEARCH is a complex project that touches on many research areas treated by the different project partners. One important goal is to hide this complexity from the end user through a consistent and context-aware user interface based on standard HTML5, JavaScript, and CSS, with ideally no additional plug-ins like Flash required. We aim at sharing one common code base for both device classes, mobile and desktop, with the user interface getting progressively enhanced [3] the more capable the user's Web browser and connection speed are. The user interface presentation can be tailored to a specific range of output devices without changing the content itself by using CSS3 Media Queries [7]. Search engines over the years have coined a common interaction pattern: the search box. We enhance this interaction pattern by context-aware modality input toggles that create modality query tokens in the I-SEARCH search box. Figure 1 shows three example modality query tokens for *audio*, *emotion*, and *geolocation*.

3. USE CASES

In order to give the reader an idea of intended I-SEARCH usage and to motivate multimodality, we introduce three use cases and involved modalities as defined by the project.

UC1: Music Expert (Desktop, Mobile).

A music expert with access to a big music archive does research on the influence of traditional folk music on today's popular music. She inputs a *rhythm* to the I-SEARCH sys-

ICMR '12, June 5–8, Hong Kong, China
Copyright ©2012 ACM 978-1-4503-1329-2/12/06 ...$10.00.

[1] Demonstration: `http://isearch.ai.fh-erfurt.de/`

[2] Screencast: `http://youtu.be/-chzjEDcMXU`

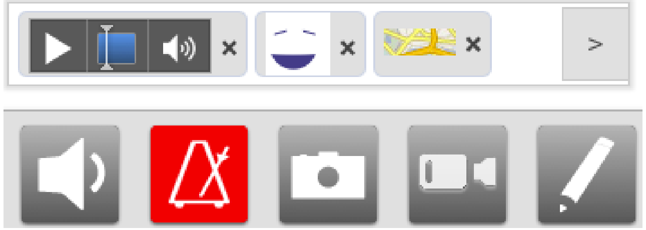

Figure 1: I-SEARCH search box with three modality query tokens for *audio*, *emotion*, and *geolocation*.

tem in order to search the archive for similar rhythm patterns. She refines her query by adding *geolocation* to limit results to a certain region and by uploading a disco *image*.

UC2: Interior Designer (Desktop).

An interior designer wants to give her client a realistic idea of office chairs. She uploads a *3D model* of a chair that almost matches her client's expectations to the I-SEARCH system, together with an *image* of the upholstery. She refines with a hand-drawn *sketch* of the chair's shape.

UC3: World Traveler (Mobile).

A traveler uses her cell phone with the I-SEARCH app to create media content with associated *geolocation* data like *videos* and *images* of the sights she visits to retrieve related content of other travelers, *text* descriptions, and *3D models* that she wants to use to map her trip on a virtual globe.

4. MODALITIES ACROSS DEVICES

In this Section, we focus on *input* modalities across *mobile* and *desktop* device classes and their support in I-SEARCH.

Audio, Image, Video We describe audio, image, and video modalities together, as they share the same interaction patterns. On desktop devices, audios, images, or videos can be uploaded from the user's hard disk via a file upload dialog or via drag 'n' drop. On some mobile devices (e.g., iOS devices) file uploading is prohibited, which is why in the longterm, as support advances, we aim at using the getUserMedia API [2]. At time of writing, access to getUserMedia is available in developer preview builds of several browsers. The fallback solution is a Flash uploader.

Rhythm On desktop devices, we support entering a rhythm by key presses or mouse tapping, whereas additionally on mobile devices a rhythm can also be captured via the device orientation API [1] by tilting the device rhythmically.

3D Object We support 3D objects on mobile and desktop via the COLLADA 3D asset exchange schema [6]. 3D objects can be inserted via a file upload dialog or drag 'n' drop.

Text On mobile and desktop, text can be entered using the keyboard or using the speech input API [10].

Emotion In order to accompany a query by basic emotional feedback, we have adapted an open-source emotion input solution [8] for mobile and desktop that transfers the slider user interface pattern to emotions from sad to happy.

Geolocation For retrieving and tracking a user's physical location, we use the HTML5 geolocation API [9], which is available in Web browsers on mobile and desktop devices.

Sketch Hand-drawn sketches can be created on mobile and desktop devices alike using a simple touch-based HTML5 canvas [5] sketch editor.

5. FUTURE WORK AND CONCLUSION

We have introduced the I-SEARCH project and three of its use cases and have shown how different input modalities are supported on mobile and desktop. Now we need to integrate the project partners' services in the back-end, in order to support multimodal output besides multimodal input.

6. ACKNOWLEDGMENTS

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

7. 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`)

8. REFERENCES

[1] S. Block and A. Popescu. DeviceOrientation Event Specification – Editor's Draft 12 July 2011. `http://dev.w3.org/geo/api/spec-source-orientation.html`.

[2] D. C. Burnett and A. Narayanan. getUserMedia: Getting access to local devices that can generate multimedia streams. `http://dev.w3.org/2011/webrtc/editor/getusermedia.html`.

[3] S. Champeon. Progressive Enhancement and the Future of Web Design. `http://www.hesketh.com/thought-leadership/our-publications/progressive-enhancement-and-future-web-design`.

[4] P. Daras, A. Axenopoulos, V. Darlagiannis, et al. Introducing a Unified Framework for Content Object Description. *Int. Journal of Multimedia Intelligence and Security (IJMIS).*, 2010.

[5] I. Hickson. HTML5 – The canvas element. `http://www.w3.org/TR/html5/the-canvas-element.html#the-canvas-element`.

[6] Khronos Group. COLLADA - 3D Asset Exchange Schema. `http://www.collada.org/`.

[7] H. Lie, T. Çelik, D. Glazman, et al. Media Queries – W3C Candidate Recommendation 27 July 2010. `http://www.w3.org/TR/css3-mediaqueries/`.

[8] G. Little. Smiley Slider. `http://glittle.org/smiley-slider/`.

[9] A. Popescu. Geolocation API Specification, 2010. `http://dev.w3.org/geo/api/spec-source.html`.

[10] S. Sampath and B. Bringert. Speech Input API Specification – Editor's Draft 18 October 2010. `http://www.w3.org/2005/Incubator/htmlspeech/2010/10/google-api-draft.html`.

[11] T. Steiner, L. Sutton, S. Spiller, et al. I-SEARCH – A Multimodal Search Engine based on Rich Unified Content Description (RUCoD). *European Projects Track at the 21st Int. World Wide Web Conf.*, 2012.