

“I-SEARCH user interfaces  
for PCs and mobile devices”

D7.3

‘’

Version: 1.0

Last Update: 20/7/2012 15:07:00 127/P7

Distribution Level: PU (Report), RE (Prototype)

* ***Distribution level***

*PU = Public,*

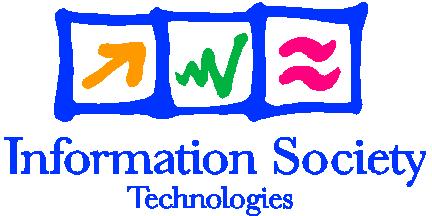
*RE = Restricted to a group of the specified Consortium,*

*PP = Restricted to other program participants (including Commission Services),*

***CO= Confidential, only for members of the I-SEARCH Consortium (including the Commission Services)***

***The research leading to these results has received funding from the European Community's Seventh Framework***

***Programme (FP7/2007-2013) under grant agreement n° 248296***

fp7.jpg

**The I-SEARCH Project Consortium groups the following organizations:**

|  |  |  |
| --- | --- | --- |
| **Partner Name** | **Short name** | **Country** |
| Centre for Research and Technology Hellas / Informatics and Telematics Institute | CERTH | GR |
| JCP-Consult | JCP | FR |
| Institut National de Recherche en Informatique | INRIA | FR |
| Athens Technology Center | ATC | GR |
| Engineering Ingegneria Informatica S.p.A. | ENG | IT |
| Google Ireland Limited | Google | IR |
| University of Genoa | UNIGE | IT |
| Exalead | EXL | FR |
| Fulda University of Applied Sciences | HSF | DE |
| Accademia Nazionale di Santa Cecilia | ANSC | IT |
| EasternGraphics GmbH | EGR | DE |

**Executive Summary:** In this deliverable, we report on work around the I-SEARCH EU (FP7 ICT STREP) project whose objective is the development of a multimodal search engine. We present the project's objectives, and detail the achieved results, amongst which a Rich Unified Content Description format.

Document Identity

|  |  |
| --- | --- |
| Title: | I-SEARCH user interfaces for PCs and mobile devices |
| Subject: | Description of user interface challenges and solutions for multimodal interaction with different classes of devices |
| Number: |  |
| File name: | WP 7.3 Google.docx |
| Registration Date: | 20 June, 2012 |
| Last Update: | 20 July 2012 |
| Authors: | GOOGLE, HSF, CERTH |

Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Version | Edition | Author(s) | Date |
| 1 | 0.1 | 1 | Thomas Steiner | 20/06/2012 |
| Comments: | Initial release | | |
| 2 |  |  |  |  |
| Comments: |  | | |
| 3 |  |  |  |  |
| Comments: |  | | |
| 4 |  |  |  |  |
| Comments: |  | | |
| 5 |  |  |  |  |
| Comments: |  | | |
| 6 |  |  |  |  |
| Comments: |  | | |
| 7 |  |  |  |  |
| Comments: |  | | |
| 8 |  |  |  |  |
| Comments: |  | | |
| 9 |  |  |  |  |
| Comments: |  | | |

**Table of Content**

1 INTRODUCTION 7

1.1 Motivation 7

1.2 Background 7

1.3 Involved Partners and Deliverable Structure 7

2 PROJECT GOALS 7

3 PROJECT RESULTS 8

3.1 Rich Unified Content Description 8

3.2 Graphical User Interface 8

3.3 Video and Image 9

3.4 Audio and Emotions 9

3.5 3D Objects 10

3.6 Visualization 10

3.7 Orchestration 10

3.8 Content Providers 10

4 SYSTEM DEMONSTRATION 10

5 Use Cases 10

6 Modalities Across Devices 11

7 RELATED WORK 12

8 FUTURE WORK AND CONCLUSION 12

9 References 12

**List of Figures**

Figure 1: I-SEARCH graphical user interface. 8

**List of Abbreviations and acronyms**

(in alphabetic order)

|  |  |
| --- | --- |
| **Abbreviation** | **Explanation** |
| API | Application Programmable Interface |
| CoFind | Collaborative Search |
| GUI | Graphical User Interface |
| HTML5 | Hypertext Mark-up Language 5 |
| pTag | Personalized Tagging Service |
| RUCoD | Rich Unified Content Description |
| UI | User Interface |
| UIIFace | Unified Interaction Interface |
| W3C | World Wide Web Consortium |
| XML | Extensible Mark-up Language |
| XMI | XML Metadata Interchange |
| SDK | Software Development Kit |
| URL | Unified Resource Identifier |
| CO | Content Object |
| DOM | Document Object Model |
| JSGF | Java Speech Grammar Format |
| REST | REpresentational State Transfer |

# INTRODUCTION

## Motivation

Since the beginning of the age of Web search engines in 1990, the search process is associated with a text input field. From the first search engine, Archie [6], to state-of-the-art search engines like WolframAlpha [1](file:///\\Zeus\Consulting\Projects\EU-I-SEARCH\Deliverables\Working%20versions\D7.3\eut13-steiner2.html#fn1x0) , this fundamental input paradigm has not changed. In a certain sense the search process has been revolutionized on mobile devices through the addition of voice input support like Apple’s Siri [1] for iOS, Google’s Voice Actions [2] for Android, and through Voice Search [3] for desktop computers. Support for the human voice as an input modality is mainly driven by shortcomings of (mobile) keyboards. One modality, text, is simply replaced by another, voice. However, what is still missing is a truly multimodal search engine. If the searched-for item is slow, sad, minor scale piano music, the best input modalities might be to just upload a short sample (“audio”) and an unhappy smiley face or a sad body expressive gesture (“emotion”). When searching for the sound of Times Square, New York, the best input modalities might be the coordinates (“geolocation”) of Times Square and a photo of a yellow cab (“image”). The outlined search scenarios are of very different nature, and even for human beings it is not easy to find the correct answer, let alone that such answer exists for each scenario. With I-SEARCH, we thus strive for a paradigm shift; away from textual keyword search, towards a more explorative multimodality-driven search experience.

## Background

It is evident that for the outlined scenarios to work, a significant investment in describing the underlying media items is necessary. Therefore, in [5], we have first introduced the concept of so-called content objects, and second, a description format named Rich Unified Content Description (RUCoD). Content objects are rich media presentations, enclosing different types of media, along with real-world information and user-related information. RUCoD provides a uniform descriptor for all types of content objects, irrespective of the underlying media and accompanying information. Due to the enormous processing costs for the description of content objects, our approach currently is not yet applicable on Web scale. We target “Company Wide Intraweb” scale rather than World Wide Web scale environments, which, however, we make accessible in a multimodal way from the World Wide Web.

## Involved Partners and Deliverable Structure

The involved partners are CERTH/ITI (Greece), JCP-Consult (France), INRIA Rocquencourt (France), ATC (Greece), Engineering Ingegneria Informatica S.p.A. (Italy), Google (Ireland), University of Genoa (Italy), Exalead (France), University of Applied Sciences Fulda (Germany), Accademia Nazionale di Santa Cecilia (Italy), and EasternGraphics (Germany). In this deliverable, we give an overview on the I-SEARCH project so far. In [Section 2](#x1-7r2), we outline the general objectives of I-SEARCH. [Section 3](#x1-8r3) highlights significant achievements. We describe the details of our system in [Section 4](#x1-75r4). Relevant related work is shown in [Section 5](#x1-78r5). We conclude with an outlook on future work and perspectives of this EU project.

# PROJECT GOALS

With the I-SEARCH project, we aim for the creation of a multimodal search engine that allows for both multimodal in- and output. Supported input modalities are audio, video, rhythm, image, 3D object, sketch, emotion, social signals, geolocation, and text. Each modality can be combined with all other modalities. The graphical user interface (GUI) of I-SEARCH is not tied to a specific class of devices, but rather dynamically adapts to the particular device constraints like varying screen sizes of desktop and mobile devices like cell phones and tablets. An important part of I-SEARCH is a Rich Unified Content Description (RUCoD) format that consists of a multi-layered structure that describes low and high level features of content and hence allows this content to be searched in a consistent way by querying RUCoD features. Through the increasing availability of location-aware capture devices such as digital cameras with GPS receivers, produced content contains exploitable real-world information that form part of RUCoD descriptions.

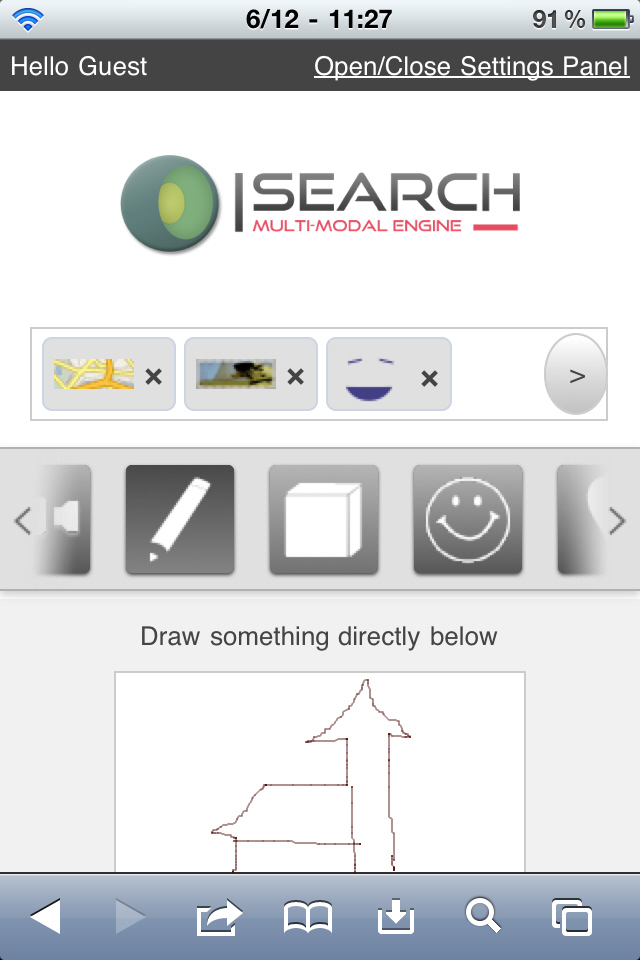
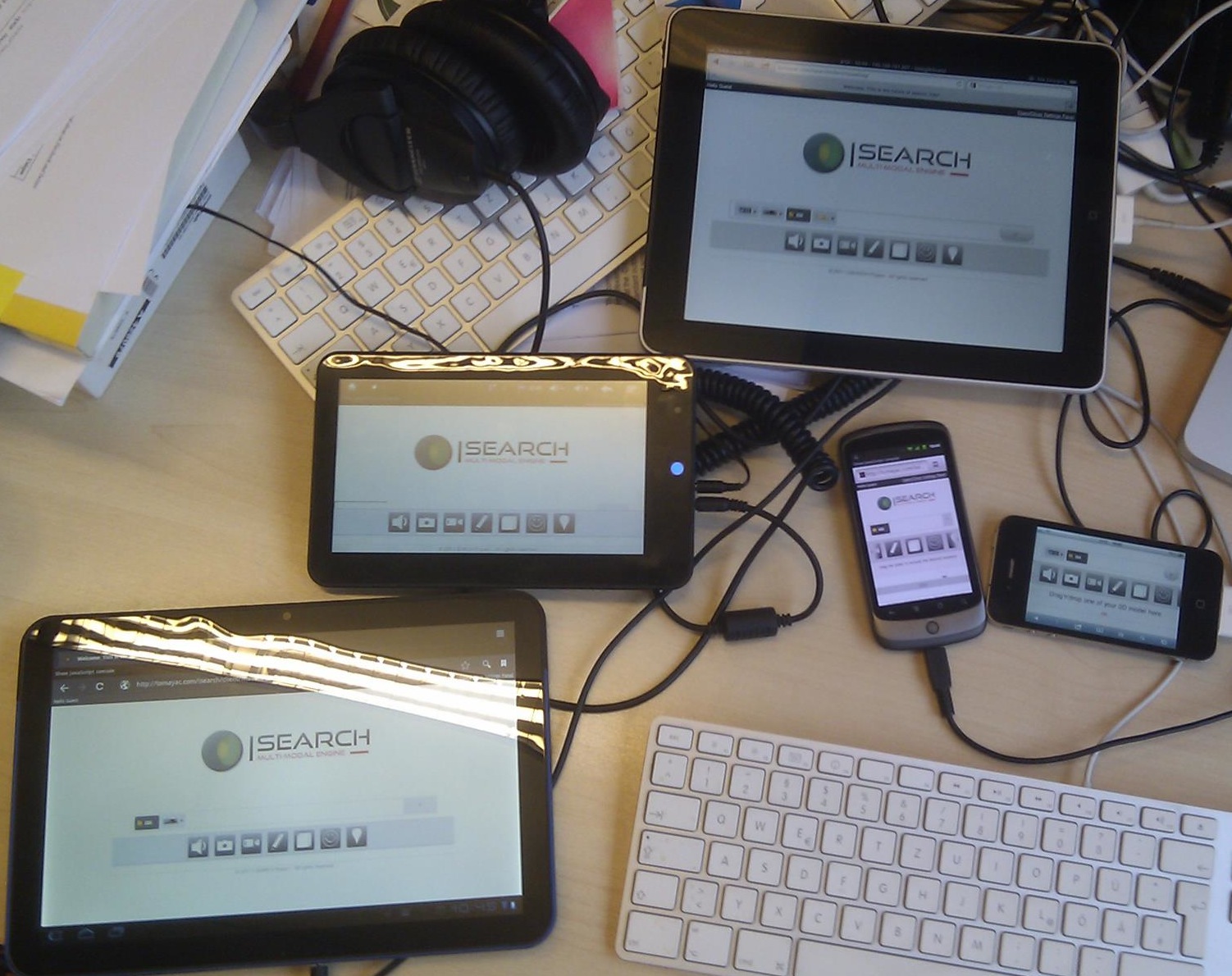
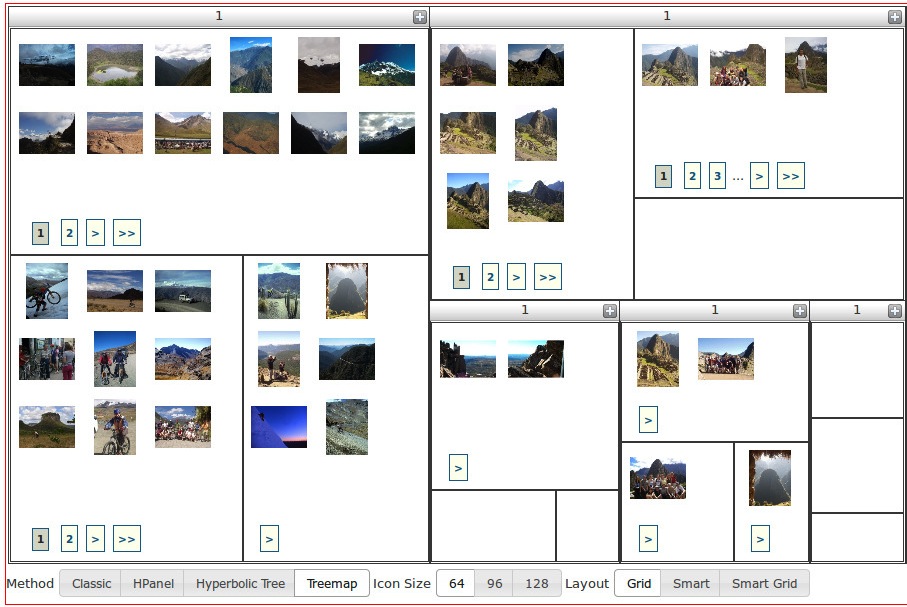
# PROJECT RESULTS

## Rich Unified Content Description

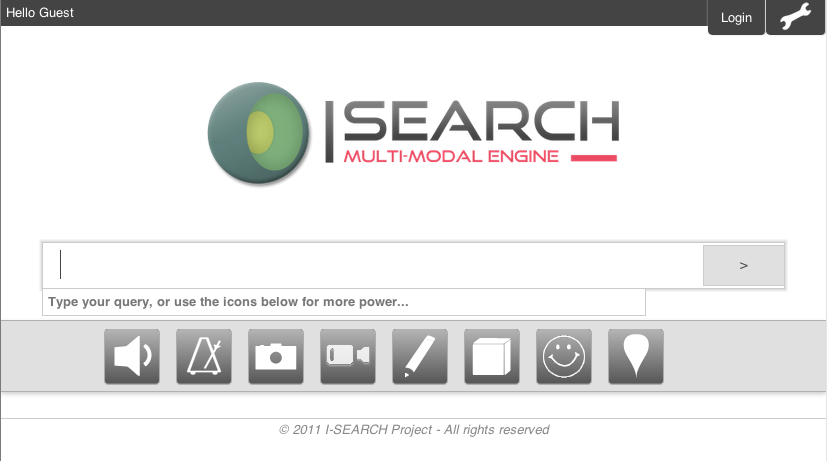
In order to describe content objects consistently, a Rich Unified Content Description (RUCoD) format was developed. The format is specified in form of XML schemas and available on the project website[2](file:///\\Zeus\Consulting\Projects\EU-I-SEARCH\Deliverables\Working%20versions\D7.3\eut13-steiner3.html#fn2x0) . The description format has been introduced in full detail in [5], [Listing 1](#x1-17r1) illustrates RUCoD with an example.

## Graphical User Interface

The I-SEARCH graphical user interface (GUI) is implemented with the objective of sharing one common code base for all possible input devices (Subfigure [1b](#x1-13r2) shows mobile devices of different screen sizes and operating systems). It uses a JavaScript-based component called UIIFace [7], which enables the user to interact with I-SEARCH via a wide range of modern input modalities like touch, gestures, or speech. The GUI also provides a WebSocket-based collaborative search tool called CoFind [7] that enables users to search collaboratively via a shared results basket, and to exchange messages throughout the search process. A third component called pTag [7] produces personalized tag recommendations to create, tag, and filter search queries and results.

  
(a) Multimodal query consisting of geolocation, video, emotion, and sketch (in progress).  
   
(b) Running on some mobile devices with different screen sizes and operating systems.   
(c) Treemap results visualization showing different clusters of images. Figure 1: I-SEARCH graphical user interface.

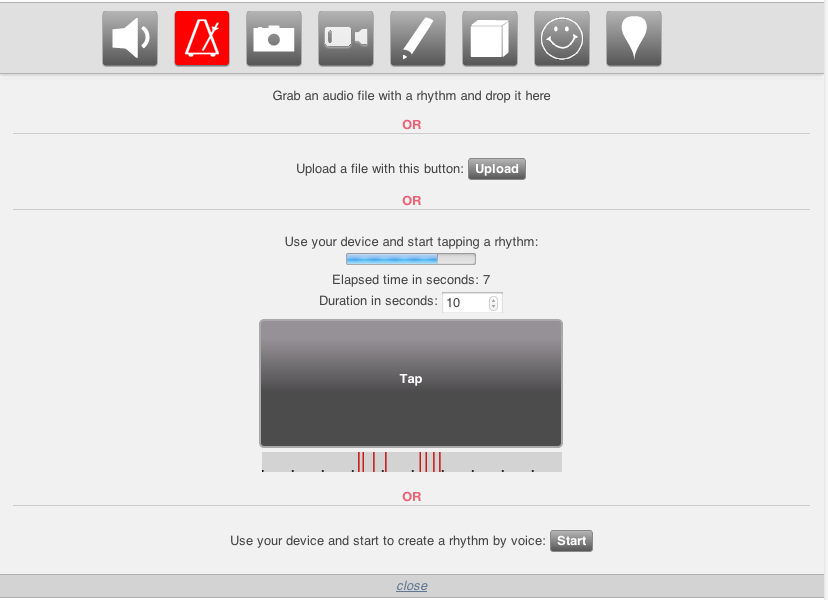
### Figure 1: I-SEARCH graphical user interface.**Standard Text Input**



### **Audio Input**

1. Drag-and-drop audio files from desktop directly into the website implemented using HTML5 File API[23]  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\1-1_sound-drag.tiff
2. It provides a classical file upload mechanisms for backwards compatibility  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\1-2_sound-upload.tiff
3. The last method allows users to make a live recording using the HTML5 getUserMedia API[15]

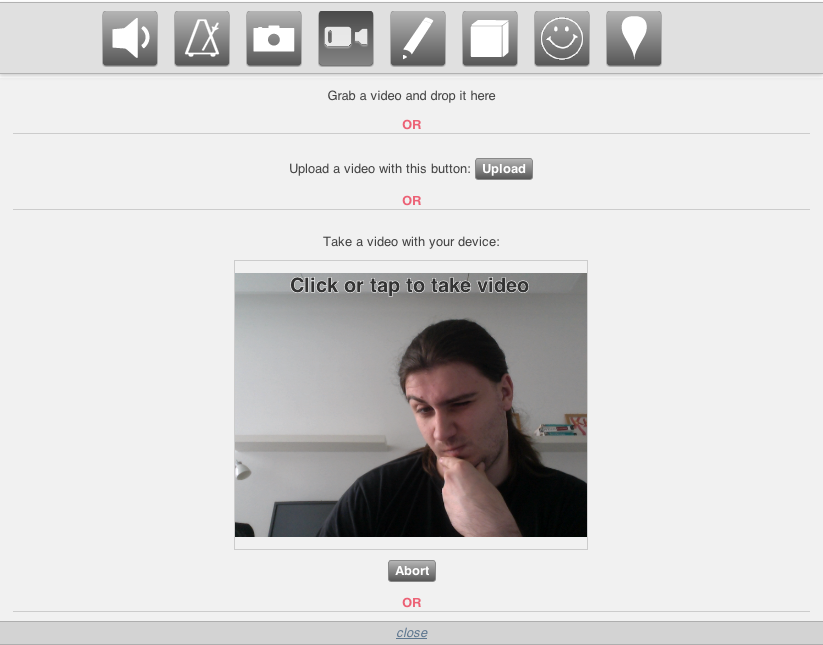
### **Rhythm Input**

1. Drag-and-drop audio files containing a rhythm from desktop directly into the designated space in the interface, implemented using HTML5 File API[23].  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\2-1_rhythm-drag.tiff
2. It provides a classical file upload mechanisms for backwards compatibility  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\2-2_rhythm-upload.tiff
3. A tap-input application is provided in which the user presses the *Tap* button in tune with the desired rhythm for ten seconds. A graphical representation of the rhythm on a scaled graph is also generated while recording using HTML5 Canvas[17] so the user can get accurate feedback of his or her actions.  
   
4. The last method allows users to make a live recording using the HTML5 getUserMedia API[15].

### **Image Input**

1. Drag-and-drop image files directly into the browser. Implemented using HTML5 File API[23].  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\3-1_picture-drag.tiff
2. It provides a classical file upload mechanisms for backwards compatibility  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\3-2_picture-upload.tiff
3. The last method allows users to take a picture on the spot using the HTML5 getUserMedia API[15] and accessing the user's webcam(in the case of a laptop/desktop), camera (in the case of a mobile device) or any other video input device available to the system

### **Video Input**

1. Drag-and-drop image files directly into the browser. Implemented using HTML5 File API[23].  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\4-1_video-drag.tiff
2. It provides a classical file upload mechanisms for backwards compatibility  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\4-2_video-upload.tiff
3. The last method allows users to record a video on the spot using the HTML5 getUserMedia API[15] and accessing the user's webcam(in the case of a laptop/desktop), camera (in the case of a mobile device) or any other video-audio input device available to the system  
   

After successfully uploading a video, the user is presented with a sequence of images (visual words) which represent symbolic distinguishable parts from the video. Depending on the length and complexity of the video this list of visual words can be of considerable size, therefore the images are displayed as a scrollable film-strip (Figure 1a). The user can navigate through the whole list, click on visual words to see the whole image and select the ones would be relevant for his or her search (Figure 1b).

1. Scrollable film-strip used for fast navigation between the various visual words  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\4-3-1_video-visualWord-select.tiff
2. Clicking on a visual word will cause it to expand to its full size for a broader view  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\4-3-2_video-visualWord-zoom.tiff

### **Sketch Input**

This menu allows users to sketch their query using HTML5 Canvas[17]. It also supports multi-touching (in the case of touch-screen or mobile devices)  
C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\5-1_draw.tiff

### **3d Object Input**

1. Drag-and-drop 3d files into the designated space in the interface, implemented using HTML5 File API[23].  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\6-1_3d-drag.tiff
2. It provides a classical file upload mechanisms for backwards compatibility.  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\6-2_3d-upload.tiff

### **Emotion Input**

I-SEARCH also provides the option of inserting an emotion to the query by dragging the emotion slider until the face expression is representative to the user's query.  
C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\7-1-2_emotion-happy.tiff

### **Location Input**

1. Automatically retrieve the user's current location from the browser using the HTML5 Geolocation API[20]  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\8-1_location-give.tiff
2. Select the desired location from an interactive Google Maps map  
   C:\Users\tak3r\Desktop\tomayac-I-SEARCH-c1037af\deliverables\images\8-2_location-map.tiff

## Video and Image

The video mining component produces a video summary as a set of recurrent image patches to give a visual representation of the video to the user. These patches can be used to refine search and/or to navigate more easily in videos or images. For this purpose, we use a technique of Letessier et al. [12] consisting of a weighted and adaptive sampling strategy aiming to select the most relevant query regions from a set of images. The images are the video key frames and a new clustering method is introduced that returns a set of suggested object-based visual queries. The image search component performs approximate vector search on either local or global image descriptors to speed up response time on large scale databases.

<RUCoD>   
  <Header>   
    <ContentObjectType>   
      Multimedia Collection   
    </ContentObjectType>   
    <ContentObjectName xml:lang=~en-US~>   
      AM General Hummer   
    </ContentObjectName>   
    <ContentObjectCreationInformation>   
      <Creator>   
        <Name>CoFetch Script</Name>   
      </Creator>   
    </ContentObjectCreationInformation>   
    <Tags>   
      <MetaTag name=~UserTag~ type=~xsd:string~>   
        Hummer   
      </MetaTag>   
    </Tags>   
    <ContentObjectTypes>   
      <MultimediaContent type=~Object3d~>   
        <FreeText>Not Mirza’s model.</FreeText>   
        <MediaName>2001 Hummer H1</MediaName>   
        <MetaTag name=~UserTag~ type=~xsd:string~>   
          Hummer   
        </MetaTag>   
        <MediaLocator>   
          <MediaUri>   
            http://sketchup.google.com/[...]   
          </MediaUri>   
          <MediaPreview>   
            http://sketchup.google.com/[...]   
          </MediaPreview>   
        </MediaLocator>   
        <MediaCreationInformation>   
          <Author>   
            <Name>ZXT</Name>   
          </Author>   
          <Licensing>   
            Google 3D Warehouse License   
          </Licensing>   
        </MediaCreationInformation>   
        <Size>1840928</Size>   
      </MultimediaContent>   
      <RealWorldInfo>   
        <MetadataUri filetype=~rwml~>   
          AM\_General\_Hummer.rwml   
        </MetadataUri>   
      </RealWorldInfo>   
    </ContentObjectTypes>   
  </Header>   
</RUCoD>

Listing 1: Sample RUCoD snippet (namespace declarations and some details removed for legibility).

## Audio and Emotions

I-SEARCH includes the extraction of expressive and emotional information conveyed by a user to build a query, and the possibility to build queries resulting from a social verbal or non-verbal interaction among a group of users. The I-SEARCH platform includes algorithms for the analysis of non-verbal expressive and emotional behavior expressed by full body gestures, for the analysis of the social behavior in a group of users (e.g., synchronization, leadership), and methods to extract real-world data.

## 3D Objects

The 3D object descriptor extractor is the component for extracting low level features from 3D objects and is invoked during the content analytics process. More specifically, it takes as input a 3D object and returns a fragment of low level descriptors fully compliant with the RUCoD format.

## Visualization

I-SEARCH uses sophisticated information visualization techniques that support not only querying information, but also browsing techniques for effectively locating relevant information. The presentation of search results is guided by analytic processes such as clustering and dimensionality reduction that are performed after the retrieval process and intend to discover relations among the data. This additional information is subsequently used to present the results to the user by means of modern information visualization techniques such as treemaps, an example of such can be seen in Subfigure [Subsection 1c](#x1-14r3). The visualization interface is able to seamlessly mix results from multiple modalities

## Orchestration

Content enrichment is an articulated process requiring the orchestration of different workflow fragments. In this context, a so-called Content Analytics Controller was developed, which is the component in charge of orchestrating the content analytics process for content object enrichment via low level description extraction. It relies on content object media and related info, handled by a RUCoD authoring tool.

## Content Providers

The first content provider in the I-SEARCH projects holds an important Italian ethnomusicology archive. The partner makes available all of its digital content to the project as well as its expertise for the development of requirements and use cases related to music. The second content provider is a software vendor for the furniture industry with a big catalogue of individually customizable pieces of furniture. Both partners are also actively involved in user testing and the overall content collection effort for the project via deployed Web services that return their results in the RUCoD format.

# SYSTEM DEMONSTRATION

With I-SEARCH being in its second year, there is now some basic functionality in place. We maintain a bleeding-edge demonstration server[3](file:///\\Zeus\Consulting\Projects\EU-I-SEARCH\Deliverables\Working%20versions\D7.3\eut13-steiner4.html#fn3x0) , and have recorded a screencast[4](file:///\\Zeus\Consulting\Projects\EU-I-SEARCH\Deliverables\Working%20versions\D7.3\eut13-steiner5.html#fn4x0) that shows some of the interaction patterns. The GUI runs on both mobile and desktop devices, and adapts dynamically to the available screen real estate, which, especially on mobile devices, can be a challenge. Supported input modalities at this point are audio, video, rhythm, image, 3D object, sketch, emotion, geolocation, and text. For emotion, an innovative emotion slider open source solution [13] was adapted to our needs. The GUI supports drag and drop user interactions and we aim for supporting low level device access for audio and video uploads. For 3D objects, we support Web GL powered 3D views of models. Text can be entered via speech input based on the WAMI toolkit [9], or via keyboard. First results can be seen upon submitting a query, and the visualization component allows to switch back and forth between different views.

# 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 system 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. .

# 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 [15]. 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 [14] by tilting the device rhythmically.

3D Object We support 3D objects on mobile and desktop via the COLLADA 3D asset exchange schema [18]. 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 [21].

Emotion In order to accompany a query by basic emotional feedback, we have adapted an open-source emotion input solution [13] 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 [20], 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 [17] sketch editor. .

# Addressing system requirements

The current status of the I-SEARCH User Interface consists of different implementations, which address multiple end user devices and needs. Either targeting a Web-based or a native mobile application, the development efforts in this task have been based on delivering the appropriate front-end, through which the I-SEARCH stakeholders can have access to the provisions of the I-SEARCH core platform and the corresponding services for content analytics, intelligent visualization through visual analytics and user interaction for relevance feedback and recommendations.

The functional and non-functional requirements in WP2 have guided the development of the whole platform and subsequently the UI modules. Thus, this section is devoted to show the way that the system functional and non-functional requirements have been considered in the UI development delivered by Task 7.3.

## Addressing the functional requirements

As per Deliverable D2.2, the functional requirements are split into eight categories, each one facilitating dedicated set of functionalities that should be provided by the I-SEARCH platform and should be reflected on the I-SEARCH front-end. However, not all these categories relate to the UI, but they could also constitute pure back office functional needs.

More specifically, the work in this task is associated with the following categories of functional requirements as reported in D2.2:

* Multimodal Query formulation
* Advanced processing of retrieved results
* Intelligent Visualisation
* User interaction
* Publish Content

In the following, we analyse the UI level functionalities, which have been implemented so far for each category of the corresponding functional requirement. For each category, we report on whether the respective functional requirements have been addressed and how. This is done both for the Web-based application and the native mobile application, through which one can have access to the I-SEARCH offerings.

### Multimodal Query formulation

One of the main objectives of the I-SEARCH project is to support different content modalities to be combined as a single query request of the end users so that the resulting content set to be as accurate as possible. During the multimodal query formulation, I-SEARCH combines the following content modalities; text, audio, video, image and 3D object. These content specific modalities can be accompanied by user generated social and expressive movements and real world information (i.e. geographical and temporal data, etc.) produced either by sensors or the end user him/herself.

The I-SEARCH target users exploit the UI functionalities to address this category of functional requirements, as reported on Table 1.

Table 1: Addressing the requirements for multimodal query formulation

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirement | Addressed in Web-based UI | Addressed in mobile app | Notes |
| Transform query to a textual format | Back office functionality | Back office functionality | A user is able to express a multimodal query, which is submitted to the I-SEARCH platform as a single textual object |
| Submit text as a query | Addressed | Addressed | A user can either free text or the tags associated with his/her profile as text queries |
| Submit audio as a query | Addressed | Addressed | A user can:   * Grab an audio file and drop it in the UI * Upload a file from local disk * Use the terminal device and start to say/sing something   For specific use cases, a user can additionally use the terminal device to tap the rhythm of a song. |
| Submit video as a query | Addressed | Addressed | A user can:   * Grab a video and drop it in the UI * Upload a video from local disk |
| Submit image as a query | Addressed | Addressed | A user can:   * Grab a picture and drop it in the UI * Upload an image from local disk   Additionally, the user can draw an image/sketch on the UI |
| Submit 3D object as a query | Addressed | Addressed | A user can:   * Grab a 3D model and drop it in the UI * Upload a 3D model from local disk |
| Submit Real World information as a query | Addressed | Addressed | A user can:   * Select to share the location from device specific information * Select the current location on the provided interactive map |
| Submit user movement as a query | Not implemented yet | Not implemented yet | Alternatively, a user can select his/her emotions from a slider |
| Combine different query options on the same query | Addressed | Addressed | A user can select multiple query choices and finally submit them by pressing a single button. |

### Advanced processing of retrieved results

Once a user query request has been performed, the I-SEARCH UI components interface with the I-SEARCH core platform, which exploits advanced data mining and presentation techniques to provide a quality of experience to the end users, with respect to the results being visualised on the graphical user interface. By consuming the relevant I-SEARCH services, the UI components are capable of projecting multidimensional clustered data structures to the end users and enabling the interaction between the end users and the multimodal and multimedia content.

Table 1 summarises on how the UI components address the functional requirements for the advanced processing of the extracted results.

Table 1: Addressing the requirements for advanced processing of retrieved results

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirement | Addressed in Web-based UI | Addressed in mobile app | Notes |
| Cluster results | Addressed | Not addressed | The system enables visualisation of the results in clusters, using different techniques |
| Scale multidimensional results for visualisation | Addressed | Addressed | The results produced from a multimodal query consist of high dimensional structures which are converted to a 2D map for visualisation |
| Detect the level of synchronisation between user movements and audio files | Back office functionality | Back office functionality |  |
| Analyse user feedback | Partially addressed | Not addressed | The UI offers the baseline functionality for filtering the results according to end user feedback |
| Support user recommendation mechanisms | Addressed | Not addressed | The UI enables end users marking the results as relevant to their query and annotating them with their own tags |

### Intelligent Visualisation

Following the RUCoD specification for the semantic representation of multimodal and multimedia content, I-SEARCH introduces visual analytics techniques to enhance the quality of results presentation to the target stakeholders through intelligent visualisation techniques. As such, the I-SEARCH UI components have been developed aiming to support the projection of different visualisation schemes and better navigate the users to the target content.

Table 1 makes an overview of the way that the UI components have addressed the functional requirements for the intelligent visualisation of the retrieved results.

Table 1: Addressing the requirements for intelligent visualisation

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirement | Addressed in Web-based UI | Addressed in mobile app | Notes |
| Visualise results on the UI | Addressed | Addressed | The users can search for the appropriate multimodal and multimedia content. The results page offers the possibility for exploring the modalities of a CO. |
| Invoke different visualisation schemes | Addressed | Not addressed | The users can select different visualisation schemes in order to browse the results set (i.e. list, hyperbolic tree, etc.) |
| Calculate similarity between items in the result set | Back office functionality | Back office functionality | The similarity of COs is exploited in order to calculate the relevance of the COs between each other and produce the list of results from a multimodal query |
| Support scoring and ranking, through both user and system controlled mechanisms | Not implemented yet | Not implemented yet |  |

### User Interaction

One of the core dimensions of the I-SEARCH objectives is the active user involvement in the identification of the most accurate and relevant results retrieved from a multimodal query request. On the UI level, the respective components have facilitated for delivering the front-end interaction blocks that enable the users actively participating in the search process and retrieving the relevant results.

Table 1 summarises on how the UI components address the functional requirements for the user interaction with the multimodal and multimedia content.

Table 1: Addressing the requirements for user interaction

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirement | Addressed in Web-based UI | Addressed in mobile app | Notes |
| Playback the multimedia items | Addressed[[1]](#footnote-1) | Addressed1 | The UI can open the multimedia items associated with the results of a query request. |
| Enable audio recording | Not implemented yet | Not implemented yet |  |
| Enable drawing sketches | Addressed | Addressed | A user can draw an image/sketch on the UI |
| Enable capturing video | Not implemented yet | Not implemented yet | Only image capturing is currently supported on the mobile application |
| Capture rhythm | Addressed | Not implemented yet | A user can additionally use the terminal device to tap the rhythm of a song |
| Capture the emotional features of the user movements | Not implemented yet | Not implemented yet |  |
| Capture Real World information, such as location | Addressed | Addressed | A user can specify his/her current location to be attached in the performed query |
| Enable interaction with the provided results | Addressed | Addressed | A user can:   * Browse the results * View the different modalities associated with a CO   Additionally, on the Web-based application the user can:   * Mark the results as relevant * Suggest tags to be attached to a multimedia item |
| Enable accessing from smartphones | - | Addressed | Dedicated mobile applications have been developed for Android devices |
| Enable authentication | Addressed | Addressed | A user can use an OpenID to be authenticated in the I-SEARCH UI |
| Provide profiling data | Addressed | Addressed | A user can retrieve his/her profile and authorise the I-SEARCH UI to track his/her search history |

### Publish Content

The I-SEARCH platform is great benefit for content providers, which can have an intermediate system to advertise their content and add value to it. Through the concept of Content Object, I-SEARCH introduces a new way to associate the different modalities in multimedia items, while multimedia is closely associated with the personal features of the end users to provide a user centric access to such content.

Table 1 makes an overview of the way that the UI components contribute to the content providers in publishing their multimodal and multimedia content.

Table 1: Addressing the requirements for publishing content

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirement | Addressed in Web-based UI | Addressed in mobile app | Notes |
| Upload multimedia files | Addressed | Addressed | A content provider can ingest multimedia items to the I-SEARCH platform. |
| Support context-aware content by attaching real world information, user specific emotional expressions and social behaviour information to it | Partially addressed | Partially addressed | A content provider can decide to attach real world information to the content |
| Create Content Objects consisting of related media content(s) of the same or different modality | Addressed | Not implemented yet | A content provider can associate multimedia items with each other and create COs. |

## Addressing the non-functional requirements

The development of the I-SEARCH UI components has also focused on addressing the system non-functional requirements, as they have been reported in Deliverable D2.2.

|  |  |
| --- | --- |
| Non-functional Requirement | Addressed in I-SEARCH UI Components |
| Persistence | This requirements does not refer to the UI components |
| Portability | The I-SEARCH UI is developed both on a Web-based HTML 5 environment and a native mobile application. Thus, the UI components address portability and can work on multiple end user devices without modifications. |
| Performance | The I-SEARCH UI has been implemented in standard Web technologies. The Web Server serves for supporting a number of concurrent users. |
| Scalability/Expandability | The I-SEARCH UI components have been developed in standard Web technologies, which enable both scalability in terms of the number of users that can be supported and the expansion of the UI to new functionalities. |
| Availability | The use of Web services for accessing the functionalities of the UI and the loosely coupled architecture used in the development of the I-SEARCH front-end can maximize the possibility for the availability of the I-SEARCH UI components. |
| Usability | As explained in section 7.1, the I-SEARCH UI components facilitate for usable and intuitive user interfaces. |
| Responsiveness | The UI components have been implemented in a way that all the supported functionalities, either being performed through the Web-based environment or the native mobile application, do respond within acceptable time limits |
| Interoperability/conformance with standards | The UI components servers for future maintainability as they have been developed using well known and widely used standards |
| Security | The UI components are accessible based on Open ID authentication, when private end user data is necessary to perform specific functionalities. |

# 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.

# 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 deliverable, 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 deliverable.

# 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/.

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

[15] 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.

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

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

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

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

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

[21] 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.

[22] 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.

---End of Document ---

1. For some content, a specific viewer might be needed [↑](#footnote-ref-1)