ABSTRACT
This paper addresses the design and development of MediaPick [1], an interactive multi-touch system for semantic search of multimedia contents. Our solution provides an intuitive, easy-to-use way to select concepts organized according to an ontological structure and retrieve the related contents. Users are then able to examine and organize such results through semantic or subjective criteria. As use case we considered professionals who work with huge multimedia archives (journalists, archivists, editors etc.).

Categories and Subject Descriptors
D.2.10 [Design]: Methodologies and Representation. D.2.11 [Software Architecture]: Data abstraction and Domain-specific architecture. H.5.2 [User Interfaces]: Evaluation/methodology, Graphical user interfaces (GUI), Input devices and strategies, Interaction styles, Prototyping, User-centered design. H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - Search process. H.3.5 [Information Storage and Retrieval]: Online Information Services - Web-based services.

General Terms
Design, Experimentation, Human Factors.

Keywords
HCI, tabletop, natural interaction, multi-touch, ontologies, web services, video retrieval

1. INTRODUCTION
In recent years, with the exponential growth of social networks and user generated content and the enhancement of technologies like semantic frameworks for automatic and manual annotation, videos have become the most comprehensive and widespread form of multimedia communication. To face this proliferation, advanced research systems are required. These provide the users a quick, consistent and efficient retrieval and sharing of multimedia information.

MediaPick is a system for semantic multimedia content based retrieval featuring a collaborative natural interaction application. It aims to the development of a context-aware, multiuser interface to domain specific knowledge base of multimedia data. It provides an advanced user centered interaction design, according to usability principles [2], which allows users collaboration on a tabletop [3] about specific topics that can be explored from general to specific concepts.

The main goals of the system are the enhancement of the search capabilities and the improvement of collaboration of end users with regard to a specific scenario: a large broadcast editorial staff where everyday retrieval of information within a huge amount of multimedia data can be overwhelming.

MediaPick allows professionals, expert or novice in a specific domain, to quickly find, display and arrange videos shots for the realization of broadcast services such as documentaries, news or TV programs on specific topics. MediaPick also increases system reliability: it provides a semantic ranking system that takes into account users’ ratings to improve the confidence of video items in future search on the semantic knowledge base.

MediaPick utilizes a simple flow of interaction. The users can interact with the system using gestures and intuitive manipulations on a multi-touch screen, performing ontology based queries, retrieving semantic related information, browsing through several results and selecting, evaluating and clustering video material. It offers a dynamic interactive visualization of ontologies graphs and semantic connections of concepts providing a strong understanding of a specific domain.

The system exploits the search engine Sirio [4] for the retrieval and the semantic reasoning on multimedia material. Sirio was developed for the European research project Vidivideo [5], it aims to improve access to semantic multimedia content. Sirio uses advanced multimedia analysis techniques that provide automatic recognition and annotation of concepts within videos as well as the Jena framework [6] and ontologies capabilities for semantic retrieval.

2. THE SYSTEM
MediaPick is a system that allows semantic search and organization of multimedia contents via multi-touch interaction. Users can browse concepts from an ontology structure in order to select keywords and start the video retrieval process. Afterwards they can inspect the results returned by the Sirio video search engine and organize them according to their specific purposes.

Some interviews to potential end-users has been conducted to better understand the issues in managing huge digital libraries; in particular we started a collaboration with the archivists of RAI, the italian public television network, in order to study their workflow and collect suggestions and feedbacks as well as to improve the design of our system.

Recently RAI technicians started the digitalization process of the analog archives: journalists and archivists are now able to search the digital libraries through a web-based system. This provides a
simple keyword based search on textual descriptions of the archived videos. Sometimes these descriptions are not very detailed or very relevant to the video content, thus making the document difficult to find. The cognitive load required for an effective use of the system often makes the journalists delegate their search activities to the archivists that could be not familiar with the specific topic and therefore could hardly choose the right search keyword. This causes the degradation of the results quality that often depends on the user’s lack of experience.

Finally, we argued that broadcast editorial staff need a simple, intuitive interface through which search, visualize and organize video results archived in huge digital libraries. MediaPick provides a solution that helps users for all these phases enabling a natural interaction approach.

2.1 Architecture

The system architecture is based on a modular approach that allows device independence in order to separate the core logic from different kind of inputs (figure 1).

MediaPick exploits a multi-touch technology chosen among various optical approaches experimented at Media Integration and Communication Center (University of Florence, Italy) [7] since 2004 [8]. Our solution uses an infrared LED array as an overlay built on top of an LCD standard screen (capable of a full-HD resolution). The multi-touch overlay is capable of detect fingers and objects on its surface and sends information about touches using the TUIO [9] protocol at a rate of 50 packets per second.

All user interface elements and business logic components have been developed in ActionScript 3.0 programming language [10] and are based on the Rich Internet Application paradigm [11]. MediaPick architecture is therefore composed by an input manager layer that communicates through the server socket with the gesture framework and core logic. The latter is responsible of the connection to the web services and media server, as well as the rendering of the GUI elements on the screen (figure 1).

From a slightly deeper point of view, the input management module is driven by the TUIO dispatcher: this component is in charge of receiving and dispatching the TUIO messages sent by the multi-touch overlay to the gesture framework through the server socket (figure 1). This module is able to manage the events sent by the input manager, translating them into commands for the gesture framework and core logic.

2.1.1 Multi-touch input manager and gesture framework

The logic behind these kind of interfaces needs a dictionary of gestures which users are allowed to perform. It is possible to see each digital object on the surface like an active touchable area; for each active area a set of gestures is defined for the interaction, thus it is useful to link up each touch with the active area in which it is enclosed. For this reason each active area has its own set of touches and allows the gesture recognition through the interpretation of their relative behavior (figure 2).

2.1.2 Core logic and Vidivideo web services communication

The core logic (figure 1) is the controller of the application: this module is connected to the Vidivideo web services using RESTful protocol and a RSS 2.0 XML as interchange format. Once the results are delivered to this module it will updates the GUI elements.

The multimedia results are streamed via RTMP by a commercial or open source media server (Red 5 or Adobe Media Server).
The core logic component is also in charge to receive information about the ontology structure in order to visualize specific domain concept graph.

2.2 Graphical user interface

The user interface adopts some common visualization principles derived from the discipline of Information Visualization [12] and is equipped with a set of interaction functionalities designed to improve the usability of the system for the end-users.

The GUI consists of a concepts view (figure 3), to select one or more keywords from an ontology structure and use them to query the digital library, and a results view (figure 4), which shows the videos returned from the database, so that the user can navigate and organize the extracted contents.

The concepts view consists of two different interactive elements: the ontology graph (figure 5), to explore the concepts and their relations, and the controller module (figure 6), to save the selected concepts and switch to the results view.

The user chooses the concepts as query from the ontology graph. Each node of the graph consists of a concept and a set of relations. The concept can be selected and then saved into the controller, while a relation can be triggered to list the related concepts, which can be expanded and selected again; then the cycle repeats. The related concepts are only shown when a precise relation is triggered, in order to minimize the number of visual elements present at the same time in the interface. After saving all the desired concepts into the controller, the user is able to change the state of the interface and go to the results view.

The results view still presents the controller, so that the user is able to select the archived concepts and launch the query against the database. The returned data are then displayed into an horizontal results list (figure 7), a new visual component placed at the top of the screen, which contains returned videos and their related concepts.

Each video element has three different states: idle, playback and information. In the idle state the video is represented by a still image and a label visualizing the concept used for the query. During the playback state the video starts playing from the frame in which the selected concept is present. A longer pressure of the video element triggers the information state, in which a panel with some metadata (related concepts, quality, duration, etc.) appears over the video.

The results list still presents the controller, so that the user is able to select the archived concepts and launch the query against the database. The returned data are then displayed into an horizontal results list (figure 7), a new visual component placed at the top of the screen, which contains returned videos and their related concepts.

Each video element has three different states: idle, playback and information. In the idle state the video is represented by a still image and a label visualizing the concept used for the query. During the playback state the video starts playing from the frame in which the selected concept is present. A longer pressure of the video element triggers the information state, in which a panel with some metadata (related concepts, quality, duration, etc.) appears over the video.
At the bottom of the results list there are all the concepts related to the video results. By selecting one or more of these concepts, video results are filtered in order to improve the information retrieval process.

At this stage the user can select a video element from the results list and drag it outside. This action can be repeated for other videos, returned by the same or other queries. Videos placed out of the list can be moved along the screen, resized or played. A group of videos can be created by collecting two or more video elements in order to define a subset of results. Each group can be manipulated as a single element through a contextual menu: it can be expanded to show a list of its elements or released in order to ungroup the videos.

All the user interface actions mentioned above are triggered by natural gestures shown in the Table 1.

### Table 1. Gesture/Action association for the user interface.

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single tap</td>
<td>▪ Select concepts</td>
</tr>
<tr>
<td></td>
<td>▪ Trigger the controller module switch</td>
</tr>
<tr>
<td></td>
<td>▪ Select video group contextual menu options</td>
</tr>
<tr>
<td></td>
<td>▪ Video element playback</td>
</tr>
<tr>
<td>Long pressure touch</td>
<td>▪ Trigger video information state</td>
</tr>
<tr>
<td></td>
<td>▪ Show video group contextual menu</td>
</tr>
<tr>
<td>Drag</td>
<td>▪ Move video element</td>
</tr>
<tr>
<td>Two-fingers drag</td>
<td>▪ Resize video</td>
</tr>
<tr>
<td></td>
<td>▪ Group videos</td>
</tr>
</tbody>
</table>

3. CONCLUSIONS AND FUTURE WORK
In this paper we presented MediaPick, a system for search and organization of multimedia contents via multi-touch interaction. Our solution was designed to help professionals working in the fields of broadcast networks, media publishing companies and cultural institutions, in managing huge amount of multimedia data, especially videos. By collecting interviews with some of these professionals we identified common issues from which we derived some useful design guidelines. Our solution integrates an intuitive natural interaction paradigm with a semantic access to multimedia information, allowing the end-user to easily build queries, visualize results and organize large sets of data.

Future developments could include the integration of MediaPick with mobile systems which could add interesting features like the ability to share and inspect search results even remotely.

4. REFERENCES


