

# An Entity-Based Platform for the Integration of Social and Scientific Services

Ronald Chenu-Abente, María Menéndez, Fausto Giunchiglia, Antonella De Angeli  
Dept. of Information Engineering and Computer Science  
University of Trento  
Trento, Italy  
{chenu; menendez; fausto; deangeli}@disi.unitn.it

**Abstract**—Over the last few years, several services and platforms for scientific production and dissemination have appeared (e.g. scholarly search engines, digital libraries, blogs, social networks, and conference management systems). Yet, these services are really fragmented and, in some cases, lack critical mass to become useful tools for the scientific community. Furthermore, most of them have failed to create a virtual environment where users' needs and activities are supported. In this paper, we present an entity-based platform aimed at integrating papers, persons, events and the services they enable. Conferences (events) are used as a starting point for collecting a critical mass of users: upon registration attendees become part of the community and can contribute to it, discussing publications, sharing content (e.g. presentations, videos, notes and pictures), or adding additional information on their page. The paper describes the user-centered research that has led to the design and presents its architecture which was shown during the International Joint Conference on Artificial Intelligence (IJCAI) 2011.

**Keywords:** *Scientific Platform; Semantic Web; User-Centered Design; Social Networks.*

## I. INTRODUCTION

From web-based services and search to social networks, Internet-based services have greatly changed the way we access information and interact with others. This is not different in scientific community, where these new services have affected and some even integrated themselves into the scientific process. Some of them are widely accepted by the scientific community, for example: scholarly search engines, conference websites, conference series websites, submission management systems, social networks, and digital library portals.

This wide variety in the approaches and services shows that many of the functionalities and services that are considered interesting for scientific research are still somewhat fragmented and dispersed. These systems are mainly related to three research areas: *information retrieval*, *social networking*, and *conference-aid services*.

No platform exists so far, that is able to manage a large body of new and legacy papers while also aggregating useful information and services related to persons and events from a research community. To bridge the seemingly antagonistic, conventional approach and the web-enabled new technologies, we propose the creation of a community platform focused not only around the papers that contain the scientific knowledge or the persons that create it, but also around the events in which

these people meet to discuss scientific knowledge. By using these three key elements and abstracted metadata entities (i.e., a specific representation of objects in the real world) enriched with semantic technology, we aim to accomplish the consistent aggregation of previously disjoint services; like providing access to conventional papers, providing information to aid and enable visiting conference events and, offering of services that would enrich the interactions and contributions with other researchers.

In particular, we want to create new ways for the members and interested people working in a research community to interact; before, during and after their conferences. To serve as a base to these interactions, we want not only to obtain, format and manage a body of legacy and new papers related to this community but also to aggregate (previously dispersed over several sites) useful information about people and events to the environment of a community platform. This platform, which was previewed during the International Joint Conference on Artificial Intelligence (IJCAI) 2011 and is still under development, would allow the members of the community to discuss their work and to share their content (e.g. presentations, videos, notes and pictures).

In this paper Section 2 explores existing tools and approaches currently used to aid the scientific process, Section 3 describes user requirements and conceptual design validation, Section 4 briefly introduces the platform developed according to the previous, and Section 5 presents the conclusions and possible extensions of this work.

## II. STATE OF THE ART

Several services and platforms have independently targeted different areas of scientific production and dissemination. Citations and bibliography databases such as Web of Knowledge<sup>1</sup>, SciVerse Scopus<sup>2</sup>, CiteSeerX<sup>3</sup>, DBLP<sup>4</sup>, ArXiv<sup>5</sup> and PubMed<sup>6</sup> aim at providing access to scholar publications stored in different repositories. Most of them provide access to multidisciplinary databases in sciences, social sciences, arts, and humanities. Some are targeted to a specific group of the

---

<sup>1</sup> wokinfo.com

<sup>2</sup> www.scopus.com

<sup>3</sup> citeseerx.ist.psu.edu

<sup>4</sup> www.informatik.uni-trier.de/~ley/db/

<sup>5</sup> http://arxiv.org/

<sup>6</sup> www.ncbi.nlm.nih.gov/pubmed/

research community such as DBPL, which provides access to academic publications in Computer Science and Information Technology, and PubMed, which focus on life sciences and biomedical topics. Web of Knowledge and SciVerse Scopus are available online only by subscription. Scholarly search engines such as Google Scholar<sup>7</sup> and Microsoft Academic Search<sup>8</sup> are alternatives to these databases. Several studies have explored differences among some of these systems [1][2][3][4]. Citation count, scholars ranking, update frequency, and content coverage are some of the variables investigated in the studies. Results vary across studies and disciplines. Some studies suggest that Google Scholar has a higher coverage of proceedings and journals [2][3], however shortcomings such as lack of reliable advanced search functionality, lack of controlled vocabulary, and issues regarding scope of coverage, update frequency [4] and quality control [3] might hinder the adoption by the entire academic community.

Many publications accessed via databases and search engines are stored in digital libraries such as ACM DL<sup>9</sup>, IEEEExplore<sup>10</sup>, and SpringerLink<sup>11</sup>. Abstracts might be free of charge but full text articles are usually subject to payment or subscription.

In the context of academic events (e.g., conferences, workshops), platforms such as conference management systems (e.g., Easychair<sup>12</sup>, Precision Conference<sup>13</sup>, Microsoft's Academic Conference Management System<sup>14</sup>), conference websites and conference series websites are broadly used. There is a considerable literature corpus on conference management systems [5][6][7][8]. Most of them target the development of functionalities such as paper submission management, peer review assignment, and review deliberation. Collaborative systems for physical conference settings are less spread. *Meme tag* was one of the first collaborative systems to be used in an event [9]. Developed by the MIT in 1997 the system consisted of wearable devices where users could write, share, subscribe and like memes/ideas, and a public display where memes were visualized and sorted by popularity. Similar systems aiming at enhancing collaboration in events and conferences have been later developed [10][11][12][13] but none of them has been fully adopted by the research community. Although preliminary results on usage and interaction among users are encouraging, user evaluations suggest that community interaction and real settings issues should be further investigated [12].

The development of platforms to support social interaction in professional environments has been a trend topic in the last years. Large companies such as IBM [14][15] and HP [16] are investing in design, development, and evaluation of internal social platforms. Share information, find experts, and create connections are among the main drivers for joining and contributing to professional networks [14][17]. The open and

collaborative nature of the research community has probably contributed to the creation of similar public social platforms such as Academia.edu<sup>15</sup>, Researchgate.net<sup>16</sup>, and Iamresearcher.com<sup>17</sup>. In fact, more than twenty public social platforms for researchers have been launched only in the last decade [18]. Academia.edu and ResearchGate.net are among the most popular platforms with over 1.5 million registered users by 9th July 2012. Furthermore, Alexa's<sup>18</sup> ranking indicates that these platforms are becoming more and more popular (as seen in Fig. 1). However, in-deep analysis on more than 30.000 user profiles in Academia.edu suggests that the level of engagement and contribution is low [18]. Some 45% of the analyzed users did not perform any explicit action after the login date. Explicit actions are those who leave a footprint in user's webpage such as updating user profile, following paper/user/research interest, or asking/answering questions. The number of asked/answered questions was especially low since the maximum asked number of questions by a user was 13 and the maximum number of answered questions was 11. Interaction among users is rare and tends to resemble real world hierarchies. In general, researchers in high positions affiliated to prestigious Universities in developed countries are more present and active than researchers at the beginning of their careers, affiliated in less prestigious Universities or in medium-low developed countries [18].

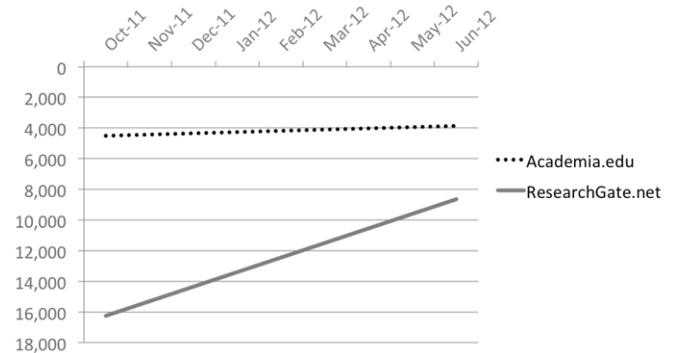


Figure 1. Evolution of Social Networks Ranking.

Although platforms such as Academia.edu and ResearchGate.net attract users' attention, they do not fully succeed at creating a common virtual place where knowledge is shared and created. Instead, results suggest that they are mainly used to host virtual portfolios. Also, no fully advantage of the democratic potential of Internet is taken and real world social hierarchies are transposed into the virtual world.

The wide variety in the approaches and services shows that many of the functionalities and services for scientific research are still dispersed. Besides the minor inconvenience of using multiple sites, the main problem is that no explicit tracking is kept over the whole process or even the interactions or discussions that are taking place. A platform used to capture and enrich the interactions that happen inside the research community has yet to appear. This platform should be able to manage a large body of new and legacy papers while

<sup>7</sup> <http://scholar.google.com/>

<sup>8</sup> <http://academic.research.microsoft.com/>

<sup>9</sup> <http://dl.acm.org/>

<sup>10</sup> [ieeexplore.ieee.org/](http://ieeexplore.ieee.org/)

<sup>11</sup> [www.springerlink.com/](http://www.springerlink.com/)

<sup>12</sup> [www.easychair.org/](http://www.easychair.org/)

<sup>13</sup> [precisionconference.com/](http://precisionconference.com/)

<sup>14</sup> [cmt.research.microsoft.com/](http://cmt.research.microsoft.com/)

<sup>15</sup> <http://academia.edu/>

<sup>16</sup> <http://www.researchgate.net/>

<sup>17</sup> <http://www.iamresearcher.com/>

<sup>18</sup> <http://www.alexa.com/>

aggregating useful information and services related to persons and events from the community.

The main part of the conceptual groundwork for the community platform presented in this paper is based on the SKO (Scientific Knowledge Object) model introduced in [19]. The creation of this still upcoming SKO model, has also taken into consideration several semantic-enabled scientific discourse models covered in [20].

### III. USER CENTERED DESIGN

A research community is the result of a complex network of services and stakeholders. Researchers, conference organizers, and institutions are some examples. Each category has different requirements. Furthermore, user requirements can be very different within same category [21]. Variables such as field of research, country of affiliation, and prestige can influence user's needs, activities, and goals. The incremental user centered design methodology adopted in this project deals with the complexity and diversity of the scenario. For the first iteration a single category of stakeholders (i.e., researchers) and a concrete design scenario (i.e., International Joint Conference in Artificial Intelligence) were selected.

#### A. User requirements

Fifteen participants were interviewed regarding different aspects of their professional life. Participants were at different stages of their research career (7 Professors, 2 Associate Professors, 3 Post-Docs, and 3 PhD students). Professional activities, in the virtual and physical world, related to the areas identified as the focus of this paper (i.e., information retrieval, social networking, and conference-aid services) were investigated.

##### 1) Information Retrieval

Participants were asked to comment on their practices when looking for information, especially for authors and publications in a specific topic. When looking for authors, the common procedure was 1) type the researcher's name and surname in a general-purpose search engine (e.g., Google and Bing) and 2) scan through the results till finding the link to the personal website. Current institution, past affiliations, resume, and list of publications were the most expected fields of information in personal websites. In general, personal websites appear to be the most trustable source of information. Most of the people preferred personal websites to digital libraries. The arguments for this preference were related to the fact that personal websites are perceived to be more complete and updated. Furthermore, websites offer a more comprehensive and complete list of publications and they are usually publicly available.

When looking for papers on a topic, interviewees had different strategies. If they knew someone working in the field, they would usually visit the author's personal website and navigate through the author's publications and referenced papers. However, when no author working in the topic was known, most of the people turned to general-purpose search engines or scholarly search engines (e.g. Google Scholar and Microsoft Academic Search). When using general-purpose search engines some users preferred visiting personal websites, even if they did not know the author. Others preferred visiting

digital libraries, such as ACM and IEEE. The arguments for visiting digital libraries were similar to those of using scholarly search engines: given a publication it was possible to navigate through referred and refereed list of publications. Less than half of the users directly looked for papers entering keywords in digital library such as ACM or IEEE. Most of the hindrances for using Digital Libraries were related to access-fees and set of publications limited to publishers. Senior researchers seemed to prefer personal websites while younger ones preferred digital libraries or research-oriented search engines. For some interviewees, especially those in the early stages of their careers, it was hard to find most of the relevant papers in a particular area because they were not familiar with semantically connected terminology (e.g., "crowdsourcing", "human computation" and "collective intelligence").

##### 2) Social Networking

Scientific Communities such as ACM and IEEE were very well known among participants. However, most of them did not use many services provided by these communities. Digital Libraries were the most used service. Participation in conferences and discount in registrations were the most common motivations to become a member of these communities. Some of the interviewees used mailing lists provided by scientific communities, although they admitted not checking most of the emails. In some cases, people were annoyed by periodic newsletters.

All the interviewees agreed that conferences are important for social networking. Young researchers and some senior researchers preferred big conferences. In general, senior researchers who preferred large conferences were also involved in the conference organization, and social networking and project funding were an important part of their professional activities. Senior researchers who were not so much involved in project funding, conference organization, or networking among institutions preferred small conferences. Either at small or large conferences, acquaintances were important in the creation of new contacts.

Most of the interviewees belonged to online social networks, but the level of engagement varied among researchers. Most of them belonged to professional social networks such as LinkedIn<sup>19</sup> or Xing<sup>20</sup>. The second most popular social network was Facebook, although most of them were not very active and used it for personal purposes. One participant had two accounts, for personal and professional matters. Just one participant, with background in sociology, belonged to diverse social networks such as Diaspora<sup>21</sup>, Quora<sup>22</sup>, Twitter<sup>23</sup>, and Identi.ca<sup>24</sup>. Low activity was mentioned as hindrance for using Diaspora. Some of them were used for professional related activities such as asking questions and writing posts.

In general interviewees were not optimistic about social networks, the main issues were related to waste of time, feeling

---

<sup>19</sup> <http://www.linkedin.com/>

<sup>20</sup> <http://www.xing.com>

<sup>21</sup> <http://joindiaspora.com>

<sup>22</sup> <http://www.quora.com>

<sup>23</sup> <http://www.twitter.com>

<sup>24</sup> <http://identi.ca>

of being under surveillance, low activity, and privacy issues. Instead, some of them used mailing lists to keep up with the community. Although some researchers did not read emails sent to the list and many were bothered by regular notifications. Face to face interactions during conferences were broadly considered the best way of social networking.

### 3) *Conference-aid services*

Some researchers checked the program before going to the conference, although many of them did it on-site. The main reasons, identified by the interviewees, for checking the program on-site were lack of preparation time and laziness. Most of them wrote down notes in the program, although some complained about the burden of carrying around heavy conference programs. Alternatively, some interviewees created a handwritten personal schedule in a piece of paper. Title, author, and topic, if available, were relevant pieces of information when deciding whether to attend to a presentation.

Many senior researchers checked who of their contacts was to attend the conference and arranged parallel meetings. Meetings were usually arranged by email or via Instant Messaging (IM) applications. The choice usually depended on the level of closeness to the contact. During the conference, most of the interviewees arranged meetings with acquaintances. When a presentation was interesting, they tried to get in contact with the speaker either approaching him/her after the presentation or sending an email. Email addresses are usually contained in the proceedings. Most the researchers exchanged business cards. For some of the interviewees business cards were not always useful. Some senior researchers get many business cards but when they come back home, it is not always easy to remember the people who gave the business cards. Remembering names and faces seem to be harder when the person who gave the card and the researcher who received it, belong to different countries. In general, Asian names and faces are harder to be remembered by Europeans.

After the conference, some researchers checked the proceedings and read papers. Those who started new contacts during the conference, sent emails to keep in touch, although it was not always an easy task. In general, they preferred reading the papers rather than looking at the video of the presentation, in case it was available. In general, all the interviewees did not give much importance to have the slides or the video of the presentation. Time and incomplete information compared to the paper were the main issues claimed by the researchers. Interviewees were more positive when the presentation was a keynote or invited talk. In general, junior researchers were more positive about being able to watch the video of presentations.

### 4) *Main requirements*

Researchers use many different platforms for scientific production. It is interesting to notice that, in spite of containing semantically related people, information, and events, these platforms are usually disconnected. Furthermore, the existing systems do not fully succeed at fulfilling users' requirements.

Data collected during the interviews suggest that researchers have well-defined practices to retrieve information, in particular to look for authors and publications. These practices usually involve the use of different systems such as search engines, digital libraries, email, and personal and

conference websites. However, the success on information retrieval usually depends on experience and knowledge. Researchers at the beginning of their career, such as PhD students, are less familiar with relationships between author-author, author-topic and topic-topic. This lack of knowledge seems to negatively influence their performance when looking for relevant papers and authors. Furthermore, inconsistencies in content (e.g., information in personal websites, paper references) and policies (e.g., access fees) among different platforms might have a negative impact on information retrieval performance.

Although existing social technologies for researchers seem to have captured the attention of many professionals, the level of engagement and interaction among users is very low. These findings are supported by the interviews since most of the researchers do not use social networks for professional purposes. For some of them, the term "social network" is only linked to platforms such as Facebook and MySpace<sup>25</sup>. Many find them a waste of time. They prefer sending emails or attending to events to create and maintain their contacts. The importance of events in researchers' professional and social life is not reflected in the design of most of the existing professional social networks. Events are the physical place where people and information come together. They act as meeting point for existing contacts and as catalyzers of new connections. New connections are usually done through acquaintances, which limits the reach to people in the neighborhood of their social network. In some cases, new connections can be also done through content (e.g., approaching a speaker after a presentation) but relating people to content is not a trivial task in the physical world.

## B. *Conceptual Design*

As we discovered during the conceptual design, the basic semantic concepts to be shared in the identified scientific production platforms are people, information, and events. Each concept can be instanced by different sub-concepts. For this first iteration, papers are chosen as instance of information, conferences as instance of events, and researchers as instance of people. Events play a major role. In our system, events are not only the place where people and information come together but also where physical and virtual space converge. Shared physical space enhanced by a virtual platform opens many possibilities for critical mass creation, social interaction, and real-time services.

Due to data availability, the selected design case was the International Joint Conference in Artificial Intelligence. A set of usually dispersed services for researchers (Search engine, Conference website, and Digital Library) was integrated in a single platform. The prototypes provided basic services for information retrieval, social networking, and conference-aid services. The digital library contained all papers published in IJCAI since 1969. Papers contained links to their author's personal profile. Authors' personal profile contained information such as affiliation, co-authors, and publications. The conference program contained all events happening during IJCAI '11 (e.g., presentations, workshops, and social events).

---

<sup>25</sup> <http://www.myspace.com>

As shown in Fig. 2, users could explore the conference program and tag favorite events on the web or mobile phone.

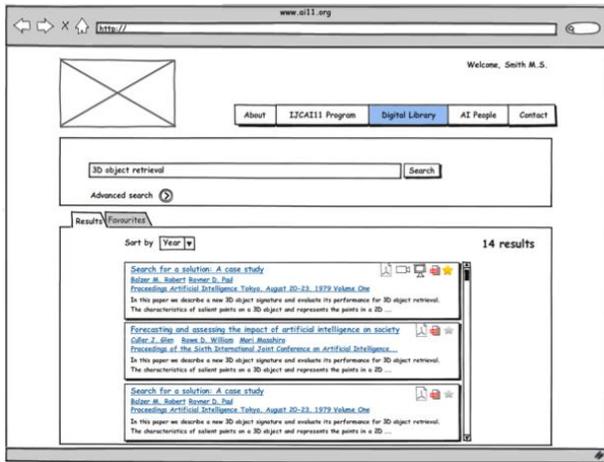


Figure 2. Mock-up for the Digital Library on the web app

### C. Concept Validation

The mock-ups were implemented and a high fidelity prototype of the system was shown during the IJCAI Conference 2011 in Barcelona for concept validation. Additionally, information on users' professional profile and suggestions for further development were gathered.

The AI Social Network was presented in the exhibitors' room during the main conference days. Three researchers, who participated in the design and development of the prototype, represented the project. The prototype was installed in a computer and a tablet and everyone who approached the booth could interact with it and provide feedback. Attendees could provide information on their professional-related activities by filling in a paper questionnaire.

#### Questionnaire data

Data from 71 attendees were collected. Most of the attendees were affiliated at Universities (76%), followed by research centers (14%) and companies (7%). Other kind of affiliations accounted for 1% of the attendees. Most of the attendees who filled in the questionnaire were from USA (17%), followed by attendees from China (11%), Spain (11%), Australia (7%), UK (7%), and Canada (5%). Regarding the position, most of the participants who filled in the questionnaire were professors (39%), followed by PhD students (25%), and researchers (14%). Most of the people who filled in the questionnaire were men (89%). Some 7% of the participants belonged to the trustee committee.

Some 83% used social networks. Although these results might be biased since participants were not randomly selected. Instead, participants voluntarily filled in the questionnaire after approaching the "AI Social Network" booth. Thus, participants might have a more positive attitude towards social networks than the average population. Most of the participants used Facebook (72%), followed by LinkedIn (65%), Academia (17%), ResearchGate (13%), and IamResearcher (10%).

TABLE I. SOCIAL NETWORKS USE BY PARTICIPANTS, TOTAL NUMBER OF USERS AND POPULARITY RANKING

Social Network	Use %	Users (Oct11)	Alexa Ranking (Dec11)
Facebook	72%	800 mill.	2
Linkedin	65%	120 mill.	13
Academia	17%	721,675	3,828
ResearchGate	13%	1,3 mill.	13,184
IamResearcher	10%	4,396	1,023,016

Facebook was the most used social network among the people who filled in the questionnaire, followed by LinkedIn, Academia, ResearchGate, and IamResearcher; Facebook was also the SN were participants consumed and produced most information, following the same SN scale as for use.

TABLE II. PARTICIPANTS' SOCIAL NETWORKS USE

	Frequency of use	Consuming	Producing
Facebook	1.96 (1.52)	0.56 (0.37)	0.19 (0.21)
Linkedin	1.07 (0.98)	0.44 (0.4)	0.16 (0.23)
Academia	0.2 (0.5)	0.08 (0.26)	0.02 (0.11)
ResearchGate	0.13 (0.34)	0.05 (0.2)	0.02 (0.12)
IamResearcher	0.11 (0.36)	0.03 (0.16)	0.02 (0.12)

#### Users' feedback

Everyone was very positive about the concept. Researchers, papers, and conferences integration in a multiplatform system was found innovative and very valuable for the community. Conference-aid services were the most popular functionalities, mainly because of the possibility of connecting conference related data (e.g., speakers and presentations) to semantic data on the virtual community (e.g., user profiles and papers).

The Digital Library was found as valuable asset. However, many people pointed out that IJCAI papers are not enough. Participants commented that papers related to the community are scattered around different repositories. Most of the participants use scholarly search engines such as Google Scholar, Microsoft Academic Search as gateway to the Digital Libraries. However, most of them confirmed limitations of these systems. The most common were related to poor categorization, lack of semantic connections, and search issues.

Although some conferences provide conference mobile applications, the presented platform provides unique functionalities connecting services offered to the temporal conference community to a permanent virtual platform. Users highly appreciated the connection between conference-aid services and the content stored in the platform. Many of them were able to come up with functionalities based on their usual activities during conferences. Many of these functionalities were targeted at meeting and finding people. Most of the people expected that the system helps them to discover people who are relevant for them, or to find experts in a given topic, department, university, or people with similar interests as them. ("This should be useful to find people that I don't know", "I would like to be able to look for an expert in a subject", "It would be good to have a list of the 5 people that are the closest to you but you don't know them yet?").

As identified in the user requirements, conferences are important for social networking. Some users were skeptical about online social networking for professional purposes. However, they liked technology to support social functionalities such as creation, maintenance, and leverage of professional connections. Popular social networks (e.g., Facebook) target population and functionalities might influence users' mental models of social networks. Young population, entertainment oriented content [22][23], and effortless cognitive interaction [24] might have negative effect in professional users' predisposition towards Social Networks, and eventually, in users' adoption. The system should not be presented as a social network but as a platform with social functionalities which enables community requirements. Access control and the possibility of creating private groups were two important issues mentioned by many of the researchers.

#### IV. SYSTEM IMPLEMENTATION

This section introduces the developed platform based on the user-centered design and as an answer to the requirements from the previous section. Starting from the platform's description of the user interface and services the direct relation with the corresponding requirements will be established, to later deepen the scope to the logic and computational infrastructure

##### A. End-User Services

End-user services are those that are directly presented to the users of the system. As such, besides offering utility, it is also important to present these services in ways that are easy to use and understand.

The analysis will be focused on three services that were identified as the most important ones in the user requirements.

##### 1) Finding and Displaying Information

As most of the researchers confirmed, the finding of the information they are looking for is mainly accomplished by searching for some property of that information (e.g., topic, author, affiliation) or by following the relations existing between them. To facilitate this process, the platform aims to offer seamless and effective search and navigation between these 'information entities', as this paper will refer to them.

Fig. 3 shows a webpage (from our live demo of the platform) displaying the list of papers presented for IJCAI11.

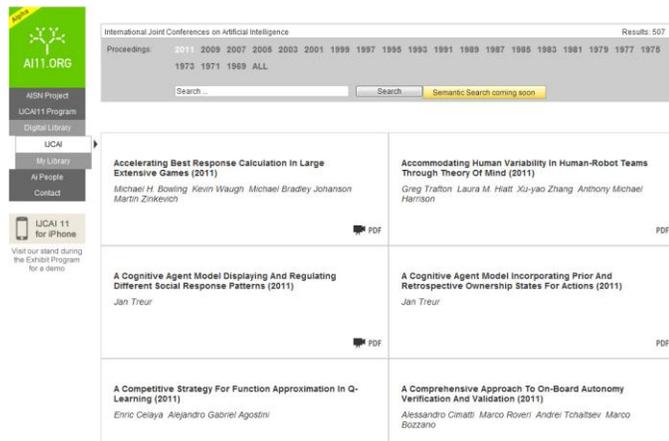


Figure 3. Paper search and browsing interface.

The screenshot from Fig. 3 shows several options to help the user find the information he is looking for:

- Semantic search bar: used to perform conventional and semantic search as described in following subsection.
- Entity Navigation: the following elements in the page can be clicked:
- Title: this would take the user to the paper profile page where all the metadata of the paper can be viewed and edited (provided the user has the permissions to do so).
- Authors: when clicked takes the user to the Author profile page.
- PDF icon: this would take the user to the actual full-text of the paper by downloading the pdf file from the IJCAI servers.
- Video icon: this would take the user to a page where she can have access to a video of the paper's presentation during IJCAI11.

Another example is shown in a person profile from Fig. 4. Note that the page not only displays the metadata related to the person (e.g. name, contact information) but also the relations that this person has with other entities in the system (e.g. co-authors, published papers). These relations are displayed as links that can be clicked to seamlessly continue the navigation throughout all the entities in the platform (as several users reported manually doing by following citations and coauthors in digital library sites).

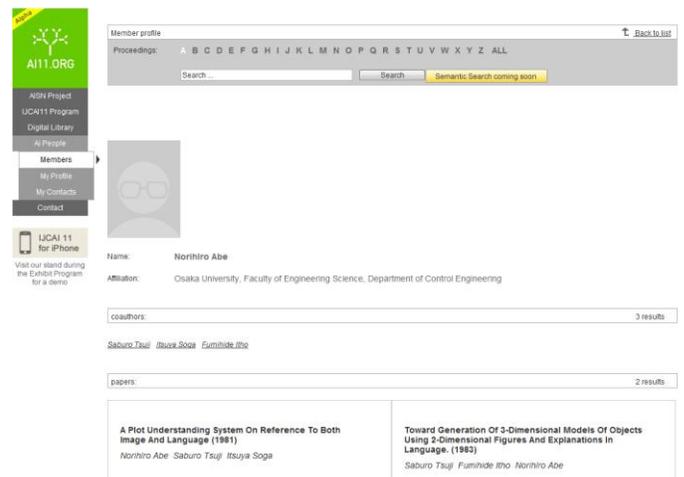


Figure 4. Person profile example.

##### 2) Favourites and Personalized Entities

As a way to assist the researchers in remembering and following their personal interests, this feature allows to define 'favourite entities' (i.e., the papers, persons and events that have a special meaning or importance for the user). As it can be noticed in Fig. 3 and Fig. 4 a small star icon appears at the top right side of each entity. Clicking on this star icon marks that entity as a favourite for the current user.

The meaning and purpose the platform gives to defining an entity as a favourite is slightly different according to the entity type:

- My Agenda (events): the purpose of this page is to show a calendar with the day and time of all the favourite events of the current user.
- My Library (papers): displays all favourite papers of the user in a single list, allowing for easy reference, reading of the pdf, or viewing of the associated video presentation.
- My Contacts (persons): shows a list of all persons marked as favourite by the current user. Future extensions include extending My Contacts page to include and capture different relations that may exist between researchers (e.g., is my collaborator, co-author, advisor).

Besides defining which entities are important for the user, the system also has plans to allow the user to add their own private annotations and tags to entities. This personalization augments the ability to define favourites by also allowing the user to capture the particular reason of interest in the target entity (e.g., A user adds a 'I need to read later' annotation to paper in its favourite list).

### 3) Mobile Support

During conferences Internet or computer access are normally limited and cumbersome. Thus, to provide in-situ assistance and services to the researchers the development of a mobile client also becomes a necessity. The mobile client developed as part of the platform mainly aims to offer key functionalities from the web client in a ubiquitous manner. More specifically, the objective is to provide support to manage the almost chaotic event of attending to a big scientific conference like IJCAI.

The mobile client is mainly focusing on managing events and helping the user build his own personalized program of the conference. Fig. 5 shows an example of how the different sub events of the conference are displayed in the mobile client.



Figure 5: Mobile client displaying the program of a conference day.

In a similar fashion to the web interface, the mobile client allows to browse different events in a conference, having access through them to papers (which can be comfortably read on the tablet version of the mobile client) and their authors. Thanks to these features the mobile client acts almost as a 'digital brochure' of the conference, assisting users to make the best of their limited conference time by allowing to quickly finding the talks and persons they are interested in.

### B. Logical Architecture

To provide the information and services necessary to enable the previously detailed user-centric services, at the logical level, the proposed platform is mainly based on three main components:

#### 1) Entity-Based Metadata Management

One of the main principles behind the proposed community platform is the use of an entity-centric abstraction to provide an uniform representation of objects, in both real and virtual world, that are relevant to the platform.

An entity  $En$ , is defined by its metadata as:  $En = \langle id, type, Attr, Rel, S \rangle$ . Where:  $id$  is a unique identifier (e.g., an URI);  $type$  is the type of entity, that is, the category to which it belongs to (e.g., the entity John is of type Person);  $Attr$  is a set of attributes composed of pairs  $attr = \langle attrname attrvalue \rangle$  describing the properties (e.g. John? date of birth is 02/01/88) of that particular entity;  $Rel$  is a set of relational attributes composed of pairs  $rel = \langle relname, relvalue \rangle$  describing the entity's relations (e.g., John is friendOf Paul) with other entities; and  $S$  is a set of services that can be leveraged on that specific entity; for example, a service "send email" can be enabled on the Person entity.

While a complete specification of each of these entity types is beyond the scope of this paper, it is important to note that these entity types are used to define the basic attributes and services a particular type of entity will have. For example, the paper entity type defines that all instances of an entity representing a paper will have the 'abstract' attribute and the 'author' relation. Furthermore, as shown in Fig. 6, all the entity types can be arranged into an entity type lattice, which allows the inheritance and extension of the metadata and services from the parent types (at top) to the children types (at the bottom).

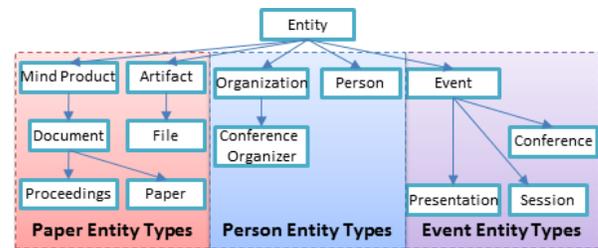


Figure 6: Entity type lattice used in the platform.

#### 2) Classification and Natural Language Processing

Classifications have been used for centuries to catalogue and search large sets of objects (e.g. classifying papers based on the topics they discuss). Classifications normally describe their contents by using natural language label (e.g. 'Computer Science', 'Databases').

The underlying idea of a semantic-based service is to have information encoded in a way that can be unambiguously interpreted by software agents, thus permitting them to find, share and integrate information more easily [26]. Unfortunately people normally annotate and classify documents by using ambiguous natural language, and trying to educate them to do otherwise requires them to go through a (normally considered) burdensome learning curve. A formal classification or a lightweight ontology ([27] and [28]) is, on the other hand, a classification where labels are written in a propositional concept language. As such, formal classification can be reasoned about far more easily than natural language sentences.

To obtain the best of the two worlds (i.e. the familiarity of natural language classification and the unambiguity of propositional concept language), approaches like [29] and [30] apply Natural Language Processing (NLP) to the natural language classifications to convert them into lightweight ontologies. This enables the platform to provide:

- Document Classification: assigning each paper to one or more categories based on their metadata. This would later be used to offer the user paper topic navigation features and also to enable better recommendation services.
- Data Integration: combining multiple sources of data is a non-trivial known issue [31] among large data and knowledge bases. This can be aided by the classification of each data source into a rooted tree and the discovery of semantic relations that exist between these trees. This allows merging multiple data sources without introducing noise (in the form of duplicates or invalid data) which is a common issue for multiple-source digital libraries.
- Semantic Search: the system can find the semantic correspondence of an object or a set of objects that corresponds to a query entered by the user if the meaning associated with the object/s is more specific or equivalent to the meaning given to the query under a common sense interpretation. This is further elaborated and exemplified below.

### 3) Semantic Search

The semantic search functionality enhances the regular search text facilities, making possible to specifically search for any of the attributes belonging to an entity (e.g. search for the paper with the following keywords, author and/or references). Furthermore, the use of a domain-specific concept knowledge base, allows the search results to contain concept-based matches besides the text-based ones.

For example, it allows the user to search for papers with the topic 'semantic search' and authors from 'Dutch universities' and to find papers about the 'concept search' approach with authors from 'Technische Universiteit of Eindhoven' (that is assuming that the underlying knowledge base contains all of the necessary concepts and relations to allow for such inference). A semantic search approach is used to perform the matching between these individual constraints and the entity attributes, i.e., it allows us to compute that the phrase 'concept search' has more specific meaning than phrase 'semantic search' and that Eindhoven is a city which is located in The

Netherlands. Furthermore, faceted search is used to specify two constraints on the paper entity, namely topic: 'semantic search' and author.affiliation: 'Dutch universities'.

Semantic search on individual attribute names and values is implemented by using the Concept Search approach [32]. Concept Search is an information retrieval approach which extends syntactic search with semantics in order to address the problems related to the ambiguity of natural language (e.g. the problems of polysemy and synonymy) by substituting words, when possible, with concepts. The main idea behind concept search approach is to reuse highly optimized retrieval models and data structures of syntactic search and preserve their efficiency while allowing for improved results when high-quality semantic information is provided. For instance, the semantic matching ([33] and [34]) of complex concepts, i.e. the core building block in the concept search approach, is implemented by using the inverted index technology.

Both features are especially useful for providing accurate results during exploratory search (allowing users to discover entities and knowledge that they did not know but that are interesting to them) and also for recommendation purposes (by performing searches with parameters related to the active user).

### C. Physical Architecture

The architecture of the physical infrastructure that supports the end-user and logical services previously discussed is shown on Fig. 7.

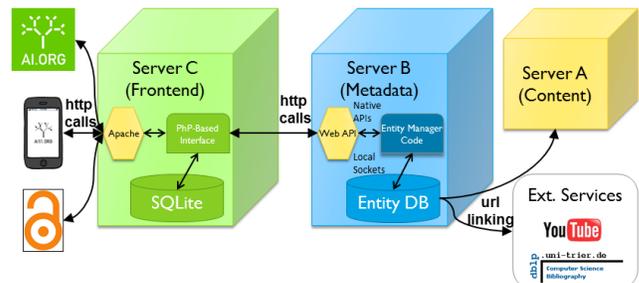


Figure 7. Platform Architecture

More specifically, in Fig. 7, the following macro-elements may be identified:

- Server A - Content: the actual content (papers, presentations, etc.) are left in their original content servers and linked through URLs. In the current version Server A is managed by the IJCAI itself, it is completely independent and external to the rest of the platform. This server is also being used at the same time to maintain the current IJCAI website<sup>26</sup> online.
- Server B - Metadata: this server implements the logical services previously discussed. The server B is considered the backend of the platform as it offers both conventional and semantic-based structures management and services for the metadata. This allows the platform to avoid any copyright or licensing complications that may arise when handling and duplicating the actual content to be discussed, while

<sup>26</sup> <http://ijcai.org/>

still being able to refer to the actual content by the use of URLs.

- External Services: using the same URL linking, materials from all over the Internet (e.g. videos, profiles on other sites) can be referenced from within the platform. This external content complements the core content from Server A and their links are added to the metadata in Server B.
- Server C: this component is mainly in charge of taking the information from the metadata server and generating the web pages/data structures that will ultimately be displayed in the clients.
- Clients: these are in charge of rendering forms for users and processing their inputs. Thanks to the underlying architecture, clients are allowed to be machines with a modest processing power (e.g. mobile devices) as the more hardware-intensive processes are carried out in the servers.

## V. CONCLUSIONS

Systems such as search engines, digital libraries, and conference websites are part of everyday technology in researchers' professional life. Users have developed strategies to search, find and assess content in these systems. However many of the reported strategies are not optimal, users need to navigate through disconnected platforms even if the content is semantically connected. Some other systems, such as social networks, have succeeded at attracting the attention of researchers but level of engagement and contribution are still an issue. Our proposal aims at tackling these issues by using a user-centered approach to design a holistic platform with metadata-enriched entities.

The developed platform is focused on four key features. First, the reformat of the vast volumes of legacy information into a reusable and easily citable format. This would allow the users to browse information from conference that happened over 30 years ago and also have access to information and services similar to the recent conferences (that were specifically tailored for the platform). Second, improved ways for helping users find conventional content, persons or events along with the context and relations that these may have (through semantic search and navigation). Specific improvements over other approaches include the ability to search for title, authors, keywords and even key concepts related to what the user wants to find and to aid the search based on the navigation of the "related items". Third, a way of introducing new web-based interactions like commenting, tagging and creation relations for all the content in the site in a certified (i.e. approved and validated by the management of the site) way. And fourth, ubiquitous/live services during conferences that, through the use of portable devices, would help attendants to find, keep track and take notes about the events happening.

With all these features integrated into the main platform, it is possible to keep track of all user queries and interactions for computing meaningful statistics about the conference and its attendants. While similar services to these already exist, they are fairly localized endeavors proving a sort of "your guide

during the conference" services. By integrating with the social network we want to also achieve "your preparation to the conference" and "this conference continues online" services that have yet to be coherently presented.

The results of the concept evaluation are encouraging. Participants were very positive about the experience and were able to come up with new services to integrate in the platform. Some of them offered technology developed within their institutions for the creation of services. User developed services open new and interesting opportunities to create a 'community created by the community'. The knowledge and skills from the members could be a valuable asset in the creation of an open community. Furthermore, this could have a positive effect on achieving critical mass and a good level of engagement. Moderation issues should be investigated.

While the live demonstration and testing of the platform at IJCAI11 was small-scale, all the different components of the system worked properly. Particularly the synchronization between the mobile client and the web client was successfully tested and it caught the attention of the participants. More realistic stress tests on the platforms are yet to be carried out.

There were, however, some limitations in this study that we would like to address in the future. First, user requirements have been collected using a group of stakeholders and part of the research community. Second, during the live demonstration of the systems we had some problems of data accuracy and duplication. Finally, no formal user evaluation of the system was performed.

We believe that this paper contributes to the research community by providing an example of integration between different research areas and the benefits of it. The design of semantic technology using a user-centered design methodology is an innovative approach. Some of the limitations have already been addressed in the period between the live tests and the publication of this paper. For example, to solve issues of data completeness and accuracy, the metadata of the platform was cross-checked and linked with DBLP and new data sources are being considered to add to the already significant amount of information in it (around 5 thousand papers and around 8 thousand authors). Furthermore, future projected features include annotation of the PDFs from legacy papers, crowdsourcing features used to improve the accuracy of the information in the platform, more social interaction features, creation of categorizations of the sub fields in the AI field (to facilitate finding papers relevant to particular subjects) and more fine-grained organizations and affiliations. (i.e., supporting the differentiation between university, department, research group).

## REFERENCES

- [1] ME. Falagas, EI. Pitsouni, GA. Malietzis, and G. Pappas, "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses", *FASEB J* 2008, 22(2)338-342.
- [2] L. Ford, and LH. O'Hara, "It's All Academic: Google Scholar, Scirus, and Windows Live Academic Search", *Journal of Library Administration*, 46(3-4), 2008, 43-52. Routledge. doi:10.1300/J111v46n03\_04.
- [3] LI. Meho, and Y. Kiduk, "Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google

- scholar”, *J. Am. Soc. Inf. Sci. Technol.* 58, 13, November 2007, 2105-2125.
- [4] IF. Aguillo, “Is Google Scholar useful for bibliometrics? A webometric analysis”, *Scientometrics* 91, 2, May 2012, 343-351. DOI=10.1007/s11192-011-0582-8.
  - [5] M. Huang, Y. Feng, and BC. Desai, “CONFYSYS: a web-based academic conference management system”. In Proceedings of the 2008 C<sup>3</sup>S<sup>2</sup>E conference (C3S2E '08). ACM, New York, NY, USA, 141-143.
  - [6] KS. Zhao, BC. Desai, R. Chen, M. Huang, and M. Lu, An online academic journal system for CINDI, In Proceedings of The Fourth International C\* Conference on Computer Science and Software Engineering (C3S2E '11). ACM, New York, NY, USA, 145-150.
  - [7] N. Di Mauro, T. Basile, S. Ferilli, “An Expert Review Assignment Component for Scientific Conference Management Systems”, In M. Ali & F. Esposito (Eds.), *Innovations in Applied Artificial Intelligence* (Vol. 3533, pp. 53-59). Springer Berlin / Heidelberg.
  - [8] M. Papagelis, D. Plexousakis, and P. Nikolaou, CONFIOUS: Managing the Electronic Submission and Reviewing Process of Scientific Conferences. In A. Ngu, M. Kitsuregawa, E. Neuhold, J.-Y. Chung, & Q. Sheng (Eds.), 2005, Vol. 3806, pp. 711-720. Springer Berlin / Heidelberg.
  - [9] R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock, “Meme tags and community mirrors: moving from conferences to collaboration”, In Proceedings of ACM conference on Computer supported cooperative work (CSCW '98). ACM, 1998, New York, NY, USA, 159-168. DOI=10.1145/289444.289490.
  - [10] FJ. McCarthy, HD Nguyen, AM Rashid, and S Soroczak, Proactive displays & the experience UbiComp project, *SIGGROUP Bull.* 23, 3, 2002, 38-41. DOI=10.1145/990017.990025.
  - [11] S. Counts, and J. Geraci, Incorporating physical co-presence at events into digital social networking. In CHI '05 extended abstracts on Human factors in computing systems, CHI EA '05, ACM, New York, NY, USA, 1308-1311. DOI=10.1145/1056808.1056903.
  - [12] T. Hope, M. Hamasaki, Y. Matsuo, Y. Nakamura, N. Fujimura, and T. Nishimura, “Doing community: co-construction of meaning and use with interactive information kiosks”, In Proceedings of the 8th international conference on Ubiquitous Computing (UbiComp'06), Paul Dourish and Adrian Friday (Eds.). Springer-Verlag, Berlin, Heidelberg, 387-403. DOI=10.1007/11853565\_23.
  - [13] K. Oh and Y. Suk Lee. “‘Who should I talk to?’: fostering a spontaneous F2F interaction in a conference”, In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion (CSCW '12), ACM, New York, NY, USA, 187-190. DOI=10.1145/2141512.2141574.
  - [14] S. Farrell, T. Lau, S. Nusser, E. Wilcox, and M. Muller, Socially augmenting employee profiles with people-tagging. In Proceedings of the 20th annual ACM symposium on User interface software and technology (UIST '07). ACM, New York, NY, USA, 2007, 91-100.
  - [15] C. Steinfield, JM DiMicco, NB Ellison, and C Lampe, “Bowling Online: Social Networking and Social Capital within the Organization”. Proceedings of the Fourth Communities and Technologies Conference, 2009.
  - [16] MJ. Brzozowski, T. Sandholm, and T. Hogg, Effects of Feedback and Peer Pressure on Contributions to Enterprise Social Media, In Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09). ACM, New York, NY, USA, 2009, 61-70.
  - [17] A. Wu, JM. DiMicco, and DR. Millen, Detecting Professional versus Personal Closeness Using an Enterprise Social Network Site. CHI 2010. doi:10.1145/1753326.1753622
  - [18] M. Menendez, A. de Angeli, and Z. Menestrina, Exploring the Virtual Space of Academia, Published in “From Research to Practice in the Design of Cooperative Systems: Results and Open Challenges”, 2012 49-63. Springer London. DOI: 10.1007/978-1-4471-4093-1\_4.
  - [19] R. Chenu-Abente, F. Giunchiglia, and L. Cernuzzi, “From Software to Artifacts: Supporting the Current Scientific Knowledge Needs”, Tech Report from Unitn, found at <http://eprints.biblio.unitn.it/1940/1/070.pdf>, 2010.
  - [20] T. Groza, S. Handschuh, T. Clark, S. Buckingham Shum, and A. de Waard, A short survey of discourse representation models, In: Proceedings 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse. Lecture Notes in Computer Science, Springer Verlag: Berlin, Washington DC, 2009.
  - [21] Unesco Science Report 2010, The Current Status of Science around the World, <http://unesdoc.unesco.org/images/0018/001898/189883e.pdf>
  - [22] AN. Joinson, “Looking at, looking up or keeping up with people?: motives and use of facebook”. In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI 2008, ACM, New York, NY, USA, 1027-1036. DOI=10.1145/1357054.1357213
  - [23] CAC. Lampe, N. Ellison, and C. Steinfield, A familiar face(book): profile elements as signals in an online social network, In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI 2007, ACM, New York, NY, USA, 435-444. DOI=10.1145/1240624.1240695
  - [24] B. Zhong, M. Hardin, and T. Sun, “Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits”, *Comput. Hum. Behav.* 27, 3, May 2011, 1265-1271.
  - [25] M. Shult, Comparing test searches in PubMed and google scholar, *J Med Libr Assoc* 2007, 95:442-5.
  - [26] I. Zaihrayeu, L. Sun, F. Giunchiglia, W. Pan, Q. Ju, M. Chi, and X. Huang, “From Web Directories to Ontologies: Natural Language Processing Challenges”, *ISWC/ASWC*, 2007, 623-636.
  - [27] F. Giunchiglia, and I. Zaihrayeu, Lightweight Ontologies, *Encyclopedia of Database Systems*, 2009, 1613-1619.
  - [28] F. Giunchiglia, D. Biswanath, and M. Vincenzo, “Faceted Lightweight Ontologies, Conceptual Modeling. Foundations and Applications”, *Lecture Notes in Computer Science*, Volume 5600/2009, 2009, 36-51, DOI: 10.1007/978-3-642-02463-4\_3.
  - [29] F. Giunchiglia, M. Marchese and I. Zaihrayeu, Encoding Classifications into Lightweight Ontologies, *J. Data Semantics*, 2007, vol 8: 57-81.
  - [30] A. Autayeu, F. Giunchiglia, and P. Andrews: Lightweight Parsing of Classifications into Lightweight Ontologies Research and Advanced Technology for Digital Libraries, 14th European Conference, ECDL2010, Glasgow, UK, 2010.
  - [31] F. Giunchiglia, M. Yatskevich, P. Avesani, and P. Shvaiko, A Large Scale Dataset for the Evaluation of Ontology Matching Systems, *The Knowledge Engineering Review Journal (KER)*, Cambridge University Press, issue 24, number 2, 2009.
  - [32] F. Giunchiglia, U. Kharkevich, and I. Zaihrayeu, Concept search. In Proceedings of ESWC, 2009.
  - [33] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, S-Match: an Algorithm and an Implementation of Semantic Matching, in proceedings First European Semantic Web Symposium, ESWS, 2004.
  - [34] F. Giunchiglia, A. Autayeu, and J. Pane.: “S-Match: an open source framework for matching lightweight ontologies”, *IOS Press, Semantic Web Journal*, May 31 2011.