INTERNATIONAL JOURNAL OF COMPUTERS
Issue 1, Volume 5, 2011

Information Retrieval and Information Extraction in Web 2.0 environment

Nikola Vlahovic

***Abstract*—** With the rise of Web 2.0 paradigm new trends in information retrieval (IR) and information extraction (IE) can be observed. Significance of IR and IE as fundamental method of acquiring new and up-to-date information is crucial for efficient decision making. Social aspects of modern information retrieval are gaining on its importance over technical aspects. The main reason for this trend is that IR and IE services are becoming more and more widely available to end users that are not information professionals but regular users. Also new methods that rely primarily on user interaction and communication show similar success in IR and IE tasks. Web 2.0 has overall positive impact on IR and IE as it is based on a more structured data platform than the earlier web. Moreover, new tools are being developed for online IE services that make IE more accessible even to users without technical knowledge and background. The goal of this paper is to review these trends and put them into context of what improvements and potential IR and IE have to offer to knowledge engineers, information workers, but also typical Internet users.

***Keywords*—** Information extraction, Information retrieval, Web 2.0, Social bookmarking, Mashups, Folksonomies.

I. Introduction

INTERNET users have been empowered with new capabilities in both using and shaping the web content unlike any time in the history of the Web. Users can share, create and rate comments on products, services, information and different entries made by both companies and institutions but also those created by other users. The popular syntagme Web 2.0 represents this paradigm shift form passive internet users who are passive consumers of available information as managed by different information editors and publishers to an era of the web where users create data for other users bypassing the middleman. This trend brought about the need for new technical and organisational solutions that are oriented towards users that support this active role. The final result is the additional growth of available online information, as well as the problem of finding relevant information. In this paper we will take a look at what are the solutions for managing and finding relevant and reliable information in Web 2.0. We can see that by letting Internet users create web content a number of incentives have appeared to empower Internet users further in order to make their search for information more efficient. We will present the issues of Information retrieval and information extraction that are connected with the rise of Web 2.0 and their further impact on trends in Web content management. A number of tools developed for this purpose will be presented as well as new paradigms such as folkonomies that stem from employment of collective intelligence over the Web.

The paper is organised as follows: In Section II information retrieval is explained as well as its impact on current information and marketing industry. In Section III Infromation extraction is described as well as properties of content where extraction methods are used. These properties are heavily influenced by the rise of Web 2.0. Relation between Information retrieval and Information extraction will also be explained as the border between these two is becoming more and more blurred. In Section IV we will concentrate on Web 2.0 and the change in user roles from passive to active participants. Users do not wish to seek information but they want the information to come to them, so some of the available technologies that are associated with Web 2.0 will also be described. Some of the most notable Web 2.0 tools for information extraction will be presented in Section V, while in Section VI discussion is given about the properties of these tools and their potential for the users but also for the further development of Web content. Finally, conclusions are made with the indications of future trends in Information extraction and web development in Section VII.

II. Information retrieval

Finding relevant information within a large collection of documents is one of the first challenges information systems had to face. For this purpose a number of computer programs and systems were developed, but it was not until the beginning of the Web that most important developments came about. Each user tries to locate documents that can yield information that he or she requires i.e. each user tries to satisfy his or her information needs. The process of identifying, searching and acquiring the potential documents that may meet these information needs is called user retrieval process [1]. All of the retrieved documents aim at satisfying user information needs expressed in natural language. A part of computer

Manuscript received October 17, 2010. This work was supported in part by Ministry of Science, Education and Sports of the Republic of Croatia as a part of the scientific research project Information, Process and Knowledge Management Systems - 067-0362980-2484

N. Vlahovic is with the Faculty of Economics and Business at the University of Zagreb, Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia (phone:+385-1-2383-333; fax: +385-1-2335-633; e-mail: nvlahovic@efzg.hr).

science that studies the retrieval of information from a collection of written documents is called Information retrieval (IR). Therefore information retrieval (IR) can be defined as a set of methods and techniques for formulating information needs of the users in form of queries. The query is then used to select a (hopefully) relevant subset of documents from a larger set i.e. the web [6]. Most of the Information retrieval systems base their retrieval method on creating a set of index terms of keywords for each document in a collection. By analyzing the user query and indexed terms system evaluates each document in collection and estimates its relevance to the user. There is a number of levels of sophistication of methods that range from basic character recognition to advanced natural language processing. The analysis can be based on the most basic level on character matching, but it may also include morphological analysis, pattern recognition and advanced natural language recognition techniques. The more advanced methods try to simulate understanding natural language texts in order to determine the content of the document.

There is a number of challenges for information retrieval systems that are solved and implemented in contemporary search engines such as Google search or Yahoo!search: (1) employing an efficient method of describing content of documents (such as indexing) and storing this information in local database, (2) efficient matching of keywords from user with terms contained in index in order to maximize the number of retrieved relevant documents – (also described as a ratio of retrieved documents and all relevant documents – i.e. recall [1]), (3) eliminating the number of retrieved documents that are falsely identified as relevant (measured as a ratio of retrieved relevant documents and overall retrieved documents – i.e. precision [1]) and (4) updating the databases with newly published web content.

Information retrieval has enabled development of one of the most successful and lucrative web-based industries at present. Some of the most successful companies that base their products and services on information retrieval are companies that develop and maintain search engines such as Google and Yahoo. Their influence on the Web and global community is an unique phenomenon as they themselves motivated the development of a new service industry – search engine optimisation (SEO). SEO marketing companies offer the service of adjusting web sites of their customers in order to reach higher visibility in search results of most popular search engines. This type of already well established marketing approach provides the highest ROI (40% of campaigns have 500% ROI) that cannot be obtained by other types of marketing strategies [2].

III. INFORMATION EXTRACTION

Sometimes users are not interested in retrieving whole documents but part of the documents and particular data contained in documents. In this case information extraction is used. Information extraction denotes any activity which goal is to automatically identify and acquire pre-specified sorts of information or data from natural language texts, aggregate them and store them in a unified and structured database[14].

The process of information extraction is twofold: firstly, precise and robust access to particular data needs to be established and secondly gathered data is structured and stored automatically in a database. The complexity of employed methods for information extraction depends on the characteristics of source texts. The method can be rather simple and straightforward if the source is well structured. If the source of information is less structured or even plain natural language, the complexity of the extraction method becomes high as it includes natural language recognition and similar processes. Information extraction can be viewed as a method in between information retrieval and natural language processing as the goal of the IE is to extrapolate concrete facts contained within a written document and representing them in some useful form (such as record in a database). Main difference between these two approaches is that for the purpose of information extraction relevant facts of interest are specified in advance, while information retrieval tries to discover documents that may have facts of interest for the user that the user is not aware of.

Information extraction is primarily based on pattern-matching algorithms, so they rely on the structure of the information source [16]. In terms of structure content of a written document can be:

- fully structured content

Fully-structured content is a content that includes detailed description of every piece of data contained within the document. This meta-data can be used to easily access any fact in the document by a program routine. The most typical documents with fully structured content are XML documents and their variations. Some of the natural language texts can also be considered as fully structured if the content satisfies additional formal rules such as it is the case with CSV (comma separated values). These texts are therefore limited to repositories of facts and raw data. Information retrieval from a fully structured content is accurate and complete if there are no faults in the document structure. Different program algorithms can be extended with error detection and correction routines, but this makes the IE process expensive and less efficient.

- semi-structured content

Semi-structured content is a content that includes only partial description of some of the data contained within the document. This type of content can also contain explicate or implicate information about the position of the beginning of different pieces of data, but the data itself is uncoupled and expressed using natural language. Natural language texts that obey the same grammatical rule of the language can be viewed as semi-structured form an information extraction point of view. In this type of content there is a great possibility of using different terms for same types of information so usage of ontologies or repositories of synonyms is advised. Information extraction from a semi-structured content is as accurate and as complete as the ontologies used are complete and precise.

- unstructured content

If the content that is used for information extraction has no meta-data about the facts within the document i.e. if the content is natural language text information extraction is limited to usually quite exhaustive pattern-matching algorithms and advanced natural language processing algorithms with partial success rates. Natural language algorithms include text mining techniques such as artificial neural networks or advanced linguistic analysis. Information extraction from unstructured content does not guarantee accuracy or completeness of data obtained while a portion of data can remain undetected.

One of the most common methods used in information extraction today is the pattern-matching method that is based on the formal description of the structure of the text that contains explicate pieces of data. For this purpose a specifically dedicated formal language for pattern-matching is developed and used called RegEx (Regular Expression).

Even though these two processes for accessing information intertwine in many areas there are significant differences between information retrieval and information extraction. Goal of information retrieval is to search for documents with information relevant to given query. The user can then access the document and find particular information on his own. The goal of information extraction goes a step further because it tries to find particular data within the content of a document. Information retrieval finds relevant documents on the web while information extraction finds relevant information in documents [6].

Both methods together make up powerful tool for accessing and organizing web information.

IV. Web 2.0 and Its impact on IR and IE

Information level of contemporary society has elevated considerably over the past two decades [9]. The increase of information dynamics within societies on a global scale can be observed as information need less time to become recorded, published and available to an increasing number of individuals. Modern media and information and communication technology – primarily the Internet and mobile technologies – have fuelled these processes on a global scale creating the fastest globalization process in all of recorded human history. The level of information that is being exchanged in a boundary-free environment has become an advantage in most of human activities especially in business and politics but also in education and entertainment [9].

On the other hand the quantity of information is becoming more and more overwhelming. Even at the beginning of the Web in 1995 serious considerations were made about developing new tools to deal with this amount of readily available information. The main issue with a huge collection of available information on the Web is the process of identifying documents that can meet information needs of any particular Internet user. With a growing quantity of information the process of identifying relevant documents becomes time consuming and the overall use of the web inefficient.

With the introduction of Web 2.0 paradigm the importance of web search (including both information retrieval and information extraction) grows even more. Web 2.0 allowed that Internet users achieve more direct communication among themselves, reducing the role of middlemen. In order to achieve this, Web 2.0 services are intrinsically required to use highly structured data, content and procedures in order to keep the overall information well organized and useful to their users. Information retrieval and especially information extraction significantly benefit from this fact. Information extraction procedure becomes less complex if the extraction is done using fully structured information source. On the other hand the created databases with repositories of extracted information can be used online or even created on-demand. The implementation of RegEx pattern-matching is also more efficient in semi-structured information sources, and features of Web 2.0 paradigm make these complex methods readily available to end users in visual and user-friendly interfaces.

Final outcome is that by information extraction allows average internet user to personalize and customize available information resources and use information sources more efficiently while still creating new context for information and enrich the quality and content of the Web even further.

Some of the services and methods that are made available through information extraction or are benefited by information extraction are given below.

A. RSS Feeds

RSS is a collection of Web formats used for publishing updates of dynamic web sites, portals and services such as blogs entries, headlines, audio and video, and other resources, in a standardized format. The abbreviation itself can be explained in several ways [13]. First of the two most common explanations is that it stands for Really Simple Syndication due to the fact that is often used to publish updates on newsportals and blogs. Instead on the characteristics of the use, the second explanation is related to the origin and composition of the technology itself as the RDF Site Summary. RDF (Resource Description Framework) is a family of World Wide Web Consortium Specifications that were originally designed as a metadata data model. By implementing RDF as a model for describing summarized site updates the model was accepted as a general model for conceptual description or modeling information that is implemented in web resource. The specification is based on XML where all relevant information about each entry is described along with additional metadata. A set of entry descriptions are usually ordered chronologically by the date of publishing forming a feed that can be subscribed to by the end users and read by using a specialized application or web service called a RSS reader.

The advantage of RSS feeds is that it can be automatically generated during the publication of an article, and therefore it is readily available to all of the subscribed users [11]. In terms of information extraction, as it is concerned with the WC3 Specification and therefore XML it creates fully structured

information sources that is easily accessible by the automated processes of information extraction tools.

B. Folksonomies

Term folksonomy was first coined by Thomas Vander to denote a bottom-up social classification [12] that was arising with the increasing popularization of Web 2.0 services such as Flickr and Delicious among others. Folksonomy can be considered an evolutionary product of social or collaborative classification of public digital content. The classification is performed by a group of people that may share common interest over certain topic or information resource by adding metadata to publish information. The process of adding metadata describing the content is repeated by all of the users and the taxonomy involving meaning of a particular information resource evolves over time. By reviewing classified content users develop a collective understanding of each term by examining the way other users use it. Finally for each describing term (usually using tags and tagging) a folksonomy is formed that promotes useful uses for each describing term and eliminating terms that are not as useful for describing content [10].

A well established folksonomy can increase the precision of information retrieval from the repository of information sources that are classified. The main characteristics of a folksonomy is that it is always created bottom-up therefore lacking any hierarchical structure, there is public availability of tags and metadata for each classified resource, and there is also social context.

Folksonomies are used for tagging many different types of content available online, such as hyperlinks to web sources (like social bookmarking service Delicious), videos (like video sharing service YouTube), pictures (like Flickr) or even retail products in online shops (like the Amazon online store).

The advantages of folksonomies for information retrieval and information extraction relate primarily to the possibility of enhancing precision of search results that is achieved outside the retrieval process. This is because information sources are better described by metadata that is indexed through the collective intelligence of users. Also similar web services can be approached through same folksonomy so that the final results are more comparable than if they were not evaluated by overall users of the web service.

C. Mashups

A mashup in web development represents a web application, web portlet or even complete web page that combines information from different information sources, reconfigures or processes them in some way and presents them in a new personalized, customized way usually revealing new context and new facts about the retrieved information. A mashup can be considered a new view on available information [3] that would be otherwise hard to realize because of lack of connection between information resources and their original presentation format, etc. The idea of mixing different content to form new aggregated or summarized information resource was brought about by the increase of published information online and the need of Internet users to stay on top of information updates. First mashups were readily available for end users such as Havarian Information Service - Alert map that uses data from more than 200 resources about different disasters being reported in real time all over the world and show this information on a world map [4]. Most available mashups are related to mapping information, photos, searches or video production.

The technology itself uses APIs (Application programming interfaces) in order to retrieve information from different online sources [15], redistribute it in a new context unforeseen by the original owners of these information sources (mashups and internet content). This is why Web 2.0 was important for creating a more stable and structured environment that could allow for the data interchange required by mashups. Lately there is a considerable effort for creating tools that can allow end users with no particular programming skills to create their own mashups [8].

In terms of information extraction a mashup can serve as an information extraction tool that can create a list of extracted information and store it online, or even create a customized RSS feed using extracted information. Also mashup can be used for pattern matching using RegEx codes in order to extract even more detailed data from available natural language sources.

V. Web 2.0 tools based on IR and IE Techniques

Even though over the last 3 years a number of tools and development environments for information retrieval and extraction in web 2.0 environment were introduced but also discontinued or replaced (such as Popfly by Microsoft [5] or QEDWiki by IBM [7]) there are a few that are being used and are given to public use while still being successfully developed further in order to provide more functionalities at a better quality.

A. Social bookmarking service Delicious

This is one of the first social bookmarking services available online (http://www.delicious.com/). Each user can store his or hers bookmarks and page links online making them available anytime anywhere, but the most important change this service introduced was the use of tags for describing each bookmark. Tags are used to organize bookmarks and form a type of indexing system that is shared between all of the users. In this way each user contributes in creating shared repository of bookmarks. Each saved link is then re-evaluated on every use and each time a new user saves it, it gains or loses relevance for a particular tag. Also users can review all of the associated tags by everyone who saved the resource, finally generating a collective consensus over the relevance and meaning of the resource in terms of a social folksonomy.

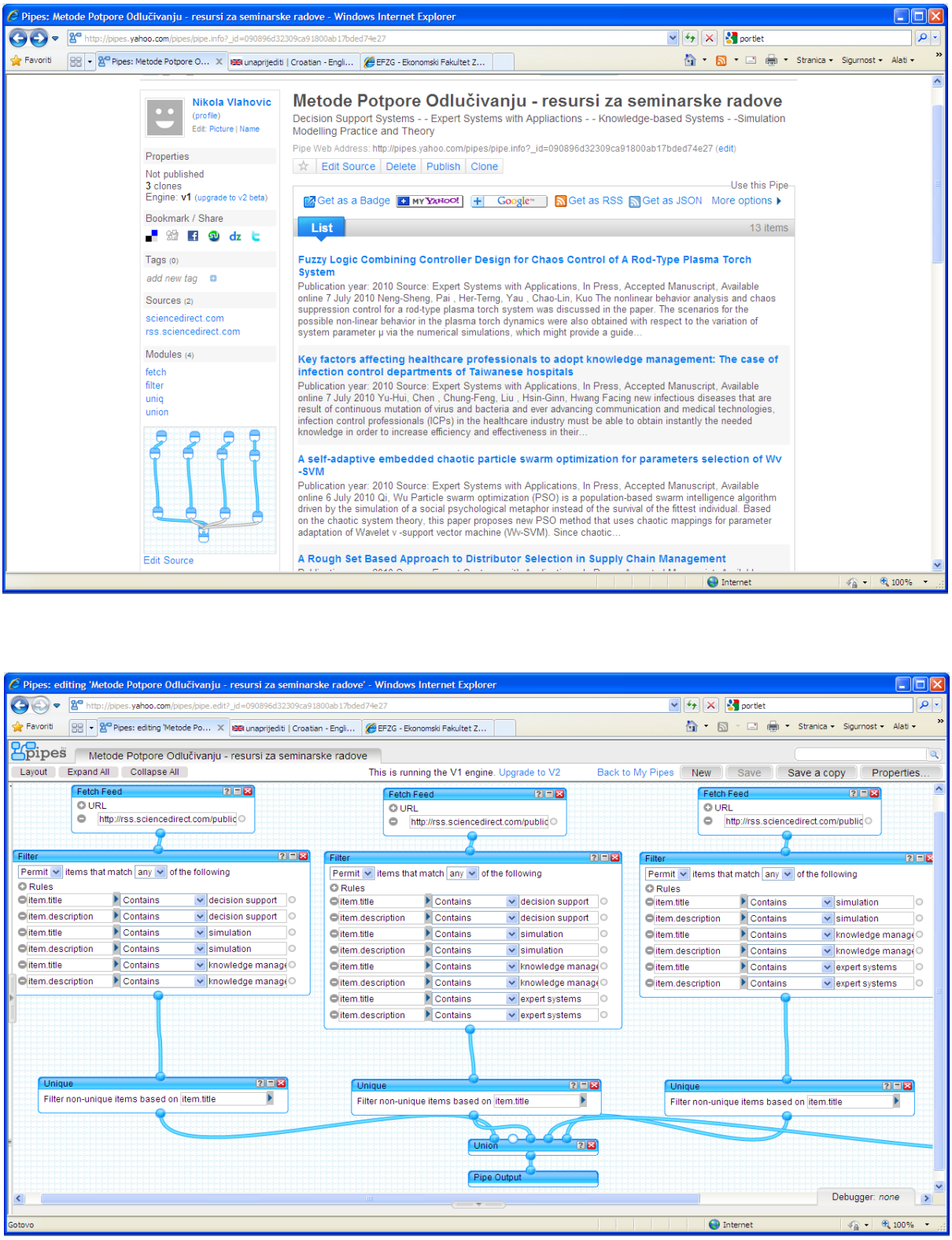

Figure 1. A list of newly published articles in three selected online journals filtered using a predefined set of key terms. A list is created using a pipe created with Yahoo!Pipes. (a) output of the pipe in end-user interface, (b) design view of the abowe pipe with its structure and keywords

This service has a high potential of identifying relevant web resources as multiple users contribute to generating and evaluating all of the acquired content, and unlike automated information retrieval algorithms or search engines it uses users common sense during these processes.

B. *Yahoo!Pipes – rewire the Web*

Yahoo!Pipes (http://pipes.yahoo.com/) is an interactive web application for manipulating and aggregating web information resources using an user-friendly graphical interface. It is used to create web mashups by creating a diagram where information is flowing through different activity elements used to analyze and process retrieved information.

This tool allows any user to employ powerful programming syntaxes in recombining and retrieving current data from any selection of internet resources, filter them and review them according to each users' specific information needs anytime anywhere.

In the upper part of fig. 1 there is a list created using Yahoo!Pipes. The list shows recent articles published in journals "Decision Support Systems", "Expert Systems with Applications", "Knowledge-based Systems" and "Simulation modeling Practice and Theory". Recent articles are also filtered according to a set of key terms relevant so that only articles that are concerned with decision support methods in business and management are shown. Selected key terms can be seen in the lower part of fig. 1. where design view of the pipe is shown. Structure and flow diagram of the pipe is also shown.

C. *Geppetto – consumer programming tool*

Geppetto (http://www.ris.fer.hr/) is a consumer-level programming environment designed for gadget composition. In this case, the building blocks used to create an application are gadgets. The Geppetto environment itself is a gadget designed to run on the iGoogle page. It is based on SOA (Service Oriented Architecture) of the Web and heavily relies on Web services and other grid capabilities. It is a graphical tool for programming and combining different available online web services that can serve a number of tasks and generate new powerful and unique active components similar to Web services.

Fig. 2. shows a blank gadget that can be edited in Geppetto interface. It consists of the starting Tab where individual gadgets can be stored and basic commands for editing (Editor), saving (Set) or canceling changes (Recovery).

The most important advantage of this service is that it allows common Internet users to create and program gadgets without having to learn a programming language.

Figure 2. A Blank Geppeto gadget

D. *Dapper – alternative representations of Web sites.*

Dapper (http://open.dapper.net/) is an interactive online service for transforming existing web pages or collections of pages into alternative forms such as RSS and other feeds, APIs, XML documents and so on. These presentation forms can then easily be used with other services such as Yahoo!Pipes or Geppetto even if the original web sites do not support news feeds or XML. Also Dapper can create a CDV data files of different resources found on web pages so it acts as an online information extraction service for downloading extracted information to local computers and databases. In this way it can easily be incorporated into existing information extraction procedures.

As with all of the described web 2.0 tools it provides simple graphical interface that empowers common internet users with information extraction capabilities. As shown in lower part of fig. 2. a process of creating a dapp consists of five steps:

1. Selection of information resources using their URL addresses
2. Collecting sample pages from each of the selected web pages. These pages should be well structured either in form of tables, ordered or unordered lists or paragraphs containing identical types of information.
3. Selection of content from each sample page. Each piece of content is stored as a separate field by highlighting part of the sample page and then previewing the selected content in the lower part of the user interface (fig. 2. b)
4. After all fields were created a complete feed can be previewed and edited.
5. The feed can be saved as a simple dapp or converted into an appropriate format such as RSS feed, JSON, YMAL, XML, CSV, HTML, XSL, or even email, iCalendar item, Google gadget API, Netvibes Module or multimedia such as Image Loop and Google Maps API.

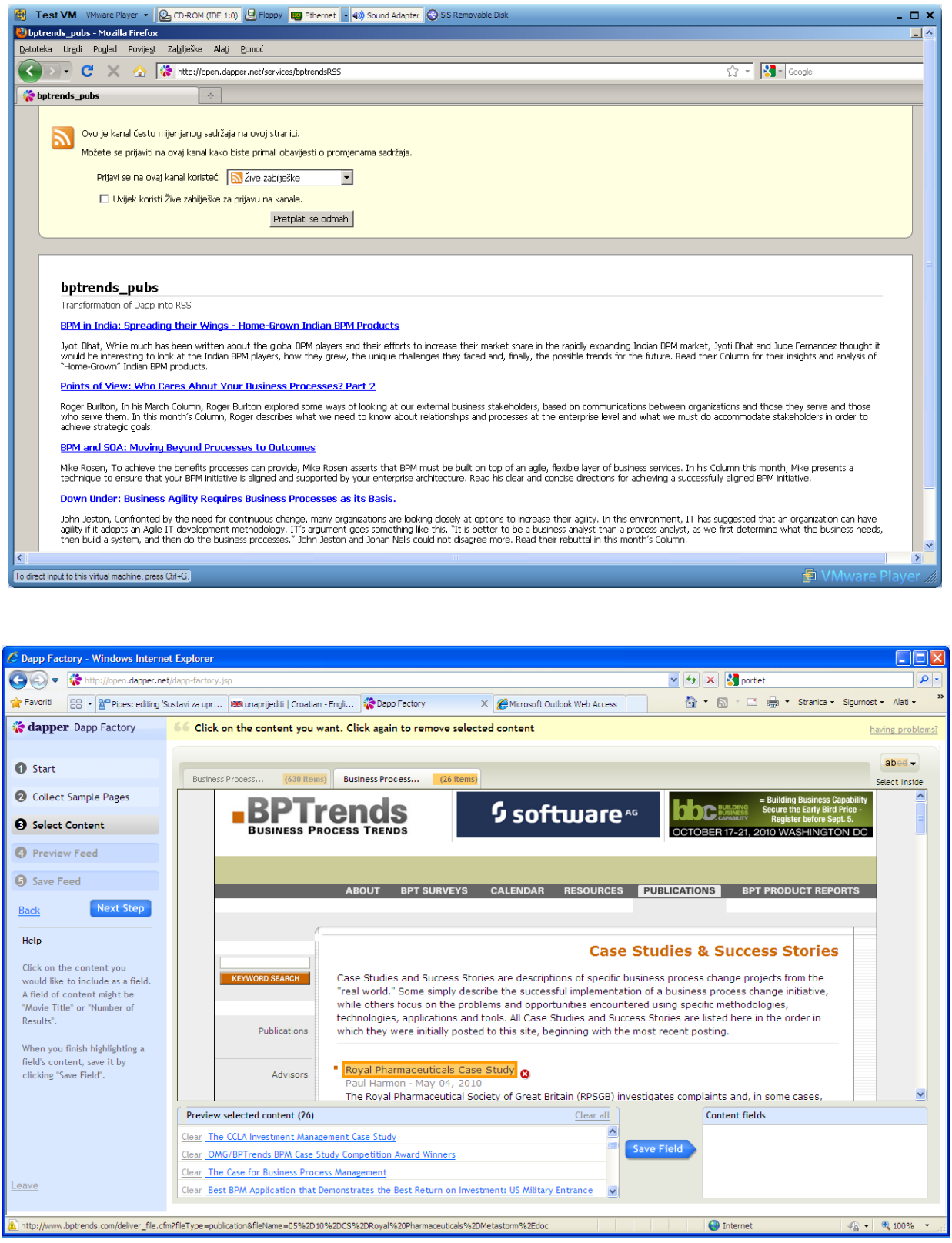

Figure 3. A dapp can be used to create RSS feeds for information resources that do not offer RSS feeds. The presented dapp is made using Dapper (http://open.dapper.net/), a) A RSS feed view of the dapp created for the site www.bptrends.com (that does not offer their own RSS feeds) as seen in Mozzila Firefox; b) Design view of the finished dapp shown above

For even greater convenience dapper is also available as an extension for the Mozzila Firefox Browser (so called DapperFox.

E. Mozzila Ubiquity – connecting the Web with language

Mozzila Ubiquity (https://wiki.mozilla.org/Labs/Ubiquity) was in development in 2008 and 2009. Its development will probably be continued at a later date. The idea is to use a unique interface for using all of the available services on-line and even combining them with the off-line and local applications, while user interface being based on natural language command lines. In this way user experience would be formed as if the user is talking with the program itself. In this way intention is to remove any border between online and offline computing as well as the borders between different applications or web services and Internet services available online. In order to achieve this natural language commands are used to communicate with the interface, while the interface itself performs all of the technical details. In terms of information retrieval and extraction, searching for information is preformed directly from the ubiquitous interface.

VI. Discussion

As we can see from the described examples, the implementation of information extraction techniques in Web 2.0 environment additionally enriched the web content. Information extraction has evolved from technical issue of information specialists to a more widely used technique for online information management both for personal and professional use. There are two main trends that can be observed.

Firstly, with the rise of Web 2.0 social aspect of information extraction becomes more pronounced. Techniques for discovering relevant information surpasses the usual methods of information retrieval and are broadened by the collective intelligence provided by folksonomies and other social bookmarking and social networking facilities. Recommendation systems of different web sites are becoming a resource for retrieval of more relevant information originating from the community of users.

Secondly, information extraction methods such as pattern-matching, recombining different information resources and creation of personalized lists and online repositories as well as dynamic updating repositories in a form of RSS feeds or downloadable CSV data files are becoming available to common internet users through simple graphical interfaces. In this way Internet users are more empowered than ever before in tasks of managing information on the internet making them more efficient and better informed individuals.

VII. Conclusions

In this paper we have presented the current developments in field of information extraction in relation to the web 2.0 paradigm. Information retrieval (IR) and information extraction (IE) are highly interrelated disciplines that have initiated a development of a new industry segment that is concerned with web search and information acquisition. The rise of search engine optimisation (SEO) services only proves the importance of these methods and their relevance in today markets and economies. On the other hand the development of Web 2.0 paradigm and introduction of new types of services and purpose of web presence that grants internet users a new degree of influence over web content while empowering them with an active role in managing the content has opened new opportunities for employment of IR an IE techniques. Goal of this paper was to present new and developing web 2.0 services that are concerned with IR and IE tasks. A number of different freely available online solutions were presented. All of the examples show that IR and IE are no longer only technical issues but that they also have a social aspect to them. Methods are not only dependent on programming algorithms but also on collective intelligence (for example in social bookmarking and creation of folksonomies) and creativity of common Internet users. In this way powerful techniques are made freely available to Internet users that can use them without having extensive programming knowledge. This fact will be an initiator of further development of trends that are still in their beginnings as the community of dedicated users of described services is rapidly growing.

References

[1] R. Baeza-Yates, B. Ribeiro-Neto, “Modern Information Retrieval, ACM Press, 1999.–64.
[2] Web Strategy Blog : 10 statistics that demonstrate the value of SEO, Available at : http://www.intraspin.com/webstrategyblog/10-statistics-that-demonstrate-the-value-of-seo [10-08-2010]
[3] E. M. Craig, “Changing paradigms: managed learning environments and Web 2.0”, Campus-Wide Information Systems, Vol. 24 No. 3, Emerald Publishing Group, 2007, pp. 152-161.
[4] RSOE Emergency and Disaster Information Services (EDIS), Havarian Information Service. Retrieved June 13, 2010 from http://hisz.rsoe.hu/alertmap/index2.php.
[5] Popfly Team : “Popfly Shutting Down”, Retrieved on June 13, 2010 from http://popflyteam.spaces.live.com/blog/cns!51018025071FD37F!336.entry?sa=197251537
[6] R. Gaizauskas, Y. Wilks, “Information extraction: Beyond Document Retrieval”, Computational Linguistics and Chinese Language Processing, Vol. 3, No. 2, August 1998, pp. 17-60.
[7] IBM AlphaWorks : QUEDWiki, Retrieved June 13, 2010 from http://services.alphaworks.ibm.com/graduated/qedwiki.html
[8] S. Peenikal, Mashups and the Enterprise: Whitepaper, MphasiS - an HP company, September 2009.
[9] J. Ridderstrale, K. Nordstrom, Funky Business: Talent makes capital dance. Pearson Education, 2000.
[10] C. Shirky, “Folksonomy”, Corante. Retrieved October 29, 2004 from http://www.corante.com/many/archives/2004/08/25/folksonomy.php
[11] C. Tzen, P. Ng, “Precisiated information retrieval for RSS feeds”, Information Management & Computer Security, Vol. 15 No. 3, Emerald Publishing Group, 2007, pp. 184-200.
[12] T. Vander Wal, You Down with Folksonomy? Retrieved December 5, 2004 from http://www.vanderwal.net/random/entrysel.php?blog=1529
[13] J. Wusteman, “RSS: The latest feed”, Library Hi Tech Vol. 22, No.4, Emerald Publishing Group, 2004, pp. 404-413.
[14] N. Vlahović, CroSE - Information management agent for Croatian capital markets, In the Proceedings of the 6th WSEAS International Conference on Distance Learning and Web Engineering, Lisbon, Portugal, September 22-24, 2006, WSEAS Press, Athens, Greece, pp. 32- 37.

[15] C. C. Tsai, C.-J. Lee, S.-M. Tang, The Web 2.0 Movement: MashUps Driven and Web Services, In the Proceedings of the 13th WSEAS International Conference on COMPUTERS, WSEAS Press, Athens, Greece, 2009, pp. 646 – 651.

[16] M. Pejic Bach, N. Vlahovic, B. Knezevic, Public Data Retrieval with Software Agents for Business Intelligence, In the Proceedings of the 5th WSEAS Int. Conf. on Applied Informatics and Communications, Malta, September 15-17, 2005, WSEAS Press, Athens, Greece, 2005, pp. 215 – 220.