3 Content Fusion Technologies

3-1 Fusion of Communication Content and Broadcast Content

MIYAMORI Hisashi, KUMAMOTO Tadahiko, NADAMOTO Akiyo, SUMI Kaoru, NAKAMURA Satoshi, MA Qiang, MINAKUCHI Mitsuru, and TANAKA Katsumi

This paper explains an overview of research results of "Fusion of Communication Content and Broadcast Content", one of the two main pillars of "Content Fusion" research project conducted at the Interactive Communication and Media Contents Group of NICT. "Fusion of Communication and Broadcast" is a conventional keyword which means technology of converging communication and broadcasting networks as an infrastructure, whereas "Fusion of Communication and Broadcast Content" represents a technology of converging Web content and TV programs at content level. Fundamental technologies and model systems were established which can efficiently utilize Internet and TV programs without complicated operations even for people who are not familiar with computer operation, such as efficient methods of accessing information and utilization methods of newly added value of information, towards the age of multitude content of TV programs and Web content available in daily lives.

Keywords
Fusion of communication and broadcast, Content fusion, Conversion and integration, Web, TV programs

1 Introduction

With recent improvements in the performance of hard disk (HD) recorders, the amount of TV program data that the average consumer can record has surged tremendously. Additionally, more web content is available than ever, thanks to the spread of high-speed Internet environments over the past few years[1].

However, there is a limit to the amount of information people can acquire in the time available. For example, searching online for information related to a TV program one is watching requires a separate series of tasks: starting a computer, accessing a search engine, and searching by keywords or similar method. To obtain information more efficiently, a framework is essential for referring back and forth among different media as needed.

Meanwhile, we are also seeing changes in the ways people use TV programs and web content. TV programs were traditionally viewed passively. Now, online content forms an “active” media that we can use as and when we wish. As a result, unprecedented forms of content consumption are becoming popular, such as:

- Watching TV in spare moments while browsing the Internet
- Watching TV and investigating various details of interest on the Internet as topics arise
- Watching TV while chatting online about the program

Taking these developments a step further, the simultaneous use of content from various media in combined formats according to the situation offers significant potential for content services of unprecedented added value.

The fusion of content described above formed the topic of a research project conducted by the Interactive Communication Media and Contents Group over the three-year period from April 2003 to March 2006. The project consisted of two parts: the fusion of communication content and broadcast content, and the fusion of digital content and real-world environments. With respect to the former, research was published on future content viewing environments for an age in which a vast amount of content is available on TV and the Internet — viewing environments that offer highly efficient means of information retrieval, new uses of information customized to the viewer, and greater added value overall.

Below, Section **2** summarizes trends and topics to date related to the fusion of communication and broadcasting, and Section **3** describes the authors' interpretation of "content fusion". Section **4** introduces examples of the development of new content viewing environments combining the Internet and television, and Section **5** presents examples of the development of content services based on the conversion of web content into other forms. Section **6** closes with a general summary.

2 Trends to date in communication-broadcast fusion

Discussions of communication-broadcast fusion first started gaining momentum around 1994. Previously, discussion had centered on the sharing of information transmission routes in the network infrastructure; that is, combining communication services (using cable TV networks) and satellite broadcasts (using communications satellites) and the like [2].

However, given the diversity both in program producers and in audiences, a significant gap developed in the pace of progress between the fields of broadcasting and communications. Traditionally, broadcast TV programs involve integrated production and distribution by broadcasters who own both the program production environment (the "soft" infrastructure) and broadcast facilities (the "hard" infrastructure), a situation that has been in place from inception to the present. In the field of communications, however, under a revision to the Broadcast Law in 1989 [3], the "soft" and "hard" infrastructure (of commissioned broadcasters) can be separated, enabling broadcast business by entities that have either program production functions or broadcast facility functions. Moreover, in 2002, the enactment of the Law Concerning Broadcast on Telecommunications Services [4] enabled broadcasting over wired networks such as ADSL, FTTH, or other broadband lines. As a result, more new broadcasters have emerged in the field of communications, and fusion services of many kinds are now offered by a variety of providers.

In the field of broadcasting, meanwhile, recent years have finally shown signs of serious efforts toward server broadcasts using metadata. It appears that server broadcasts will commence in Japan in fiscal 2007, and we can expect to see the gradual introduction of various metadata-based services — including TV-Anytime [5], P/Meta [6], and J/Meta [7].

Previous discussions of communication-broadcast fusion can be seen as focusing mainly on infrastructure sharing, diversification of providers, and program viewing supported by storage equipment.

However, very little attention has been paid to fusion specifically enabling viewers to combine program and communication content as they wish. This sort of fusion would naturally necessitate a framework for the appropriate handling of copyright; this issue aside, it is precisely this type of viewing — involving the combination of diverse content — that we define as communication-broadcast fusion.

The basic underlying technology required to implement this content fusion has been at the focus of a range of relevant research.

For example, a method of searching for web pages similar in content to TV news pro-

grams has been proposed by Henzinger and other researchers at Google, USA[8]. In this method, news programs can be treated as streams of closed-captioned information that provide the basis for a series of searches for similar web pages. Searching for complementary information[9] (described below) has been proposed as a method not only of searching for similar pages but also for complementary pages reflecting similar content from other perspectives.

In addition, TV2Web[10] and Video Manga[11] have been proposed as viewing interfaces that provide an efficient way to get an overview of programs or find scenes. TV2Web creates storyboard structures of various levels of detail, comprising thumbnails of video segments and subtitle data. Through a "zooming" feature, users can zoom in and out of the structured data to switch between elements smoothly, in an efficient approach to understanding the overall nature of programs or to finding scenes. Video Manga arranges key frames (of various sizes and determined by the relative importance of video segments) in a comic-book-like presentation format as a summary, enabling viewers to grasp important situations at a glance. Both of these approaches essentially present a reduced amount of information to users, but in the method proposed in WA-TV[12] (described below), access to information of interest is further facilitated by the integration of information related to TV programs in a screen of video segments. This approach reduces the amount of program content shown to users while increasing the amount of information offered from other perspectives.

In addition, a zooming interface[13][14] is a common means of efficiently viewing a great deal of information on a single screen. Zooming interfaces are mainly employed when users navigate through the information presented. In some instances, enlarging a portion of interest presents more detailed information; in others, the same information is presented in a variety of sizes. Although many R&D case studies on zooming interfaces have presented numerous conclusions, in short we can say that the conventional zooming function is generally dependent on the individual applications within a system. On the other hand, with respect to the concept of "zooming cross-media"[15] (described below), a zooming description language has been proposed that can describe the actions of zooming operations. Specifically, this language encodes dynamic control of the current display style, thus enabling flexible implementation of zooming functions in a common browser.

Moreover, Webstage[16] has been proposed as a method of converting web content into other media for information acquisition, in which each segment of a web page is treated as a program element and presented through sound and graphic characters. Meanwhile, wEE [17], u-Pav[18], Web2Talkshow[19], and Interactive e-Hon[20] (described below) have been explored as presentation techniques that support comprehension (through rephrasing suited to user ability) or better internalization (through methods of presentation that render the given content more familiar to the user). These methods of content generation have been proposed specifically as ways of communicating greater nuance and detail.

3 Defining fusion of communication-broadcast content

Here we summarize our interpretation of the fusion of communication and broadcast content.

As mentioned in Section **2**, much has been discussed concerning communication-broadcast fusion in terms of program distribution over broadband Internet. These discussions mainly deal with how information is streamed or distributed through the infrastructure of communication and broadcast networks. Specifically, the main focus of investigation has centered on approaches to combining infrastructural technologies of content distribution.

Although this combination is certainly a key subject, it is our view that, once this com-

munication-broadcast fusion is achieved, the approaches taken in applying this fusion technology will be equally important. How will content be combined through these approaches — specifically, web content (communications) and TV content (broadcasting)? What kinds of content services are feasible? What new value will be offered to the user? Rather than calling these approaches "communication-broadcast fusion", we refer to them as embodying "communication-broadcast content fusion". In this paper below, "communication content" is limited to web content, while "broadcast content" is restricted to TV programs in our introductions of development examples.

4 Internet-TV fusion

4.1 Complementary TV-web viewing with next-generation HD/DVD recorders — side-trip viewing

Like a stop on the way to one's destination, side-trip viewing is a feature enabling TV viewers to find the answer to any sudden questions immediately, without missing scenes of interest. After finding answers to his or her questions, the viewer can continue watching the program. Specifically, when a question comes to mind, the viewer can pause the program and perform a "zoom-in" operation. The program is displayed at reduced size and web pages with content thought to be most complementary to the scene are shown under the program window. Selecting a web page enlarges it, effectively presenting it as a regular web page. After viewers have found the information they were looking for, they can resume watching the paused scene (Fig. 1).

This approach applies a method of searching for complementary information[9] in which web pages with content complementary to the scenes are sought in real time, based on analysis of the TV program's closed-captioning titles. In searching for complementary information, special data called topic structures are extracted from the closed-captioning data as the basis for structured questions that may yield details or lead to related topics of interest concerning the captions or content. In turn, these questions are used for web searches. Unlike traditional "similarity searching", this approach enables searching on a more detailed level or from varied perspectives.

This system eliminates the need to enter search keywords, thus allowing for the efficient use of Internet-based information even by users unfamiliar with computer operations.

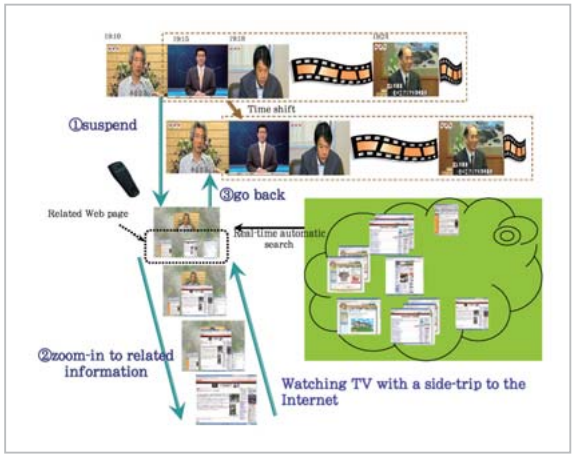


Fig.1 Side-trip viewing function

4.2 WA-TV: TV-to-web conversion and integration in related web content

"Webifying and Augmenting TV" (WA-TV) describes a system for enhancing the "browsability" of TV programs by creating easily accessible web content combining desired scenes with related information[12]. As shown in Fig. 2, WA-TV presents thumbnails of TV program scenes next to closed-caption-

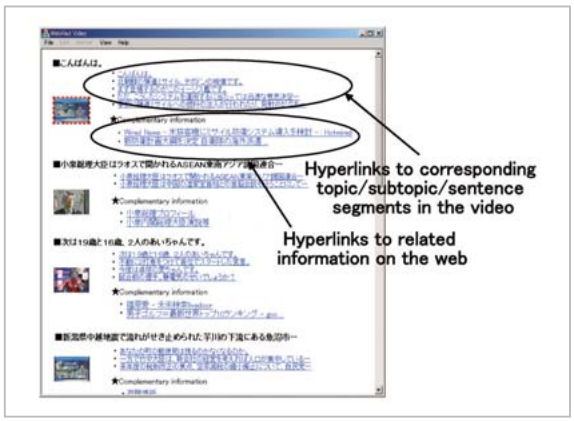


Fig.2 Sample screen from Webifying and Augmenting TV (WA-TV), which converts programs to web content and provides supporting information

ing data for the scenes. This data is converted to web content arranged in chronological order from the top to bottom. Additionally, top-ranked results from searching for complementary information are shown below the subtitle data. As with side-trip viewing, this system enables easy and efficient use of information on the Internet even by users unfamiliar with computer operations, and it offers the advantages of rendering TV programs easier to "browse" and understand.

4.3 Applying viewer reactions using live chat about programs

Using live chat with respect to ongoing programs represents a potentially effective method of obtaining relevant information[21]. In such a system, viewers of TV programs would congregate in a chat community on the Internet and enter comments regarding their impressions of the program or specific content in real-time, enjoying a form of online conversation. Changes in this chat data over time or particular expressions can be analyzed to detect the extent of viewer reactions or intensity of emotions — from excitement, enjoyment, or interest to sadness or despair. This analysis in turn would allow for the presentation of viewer reaction rankings (for example, ranking scenes from the previous week that elicited the greatest reaction or excitement among viewers, by frequency), viewing with others who have indicated similar (or different) values, or other means of program viewing offering new added value not available through traditional media[21].

4.4 Integrated search engine for recorded programs and web content

A prototype search engine has been developed that enables integrated searching not only of web content but also programs recorded on a user's computer[22]. (Although Sections **4.1** to **4.3** discuss applications for either live or recorded programs, this section deals with an application exclusively for recorded programs.) Integrated ranking of web content and recorded programs is implemented as follows. Query candidates are extracted from closed-captioning of recorded programs, and existing search engines are used to obtain web page candidates. Characteristic vectors common to the web page candidates and recorded programs stored on the user's personal video recorder (PVR) are generated using a vector space model, and an integrated index is created. Integrated rankings are generated by calculating the similarity between a query and the characteristic vectors in the integrated index. Accessing other media and controlling the level of detail is achieved by zooming, implemented through zooming cross-media[15] technology.

5 Conversion of web content to other media

5.1 Web2TV with Emotional Expression (wEE)

wEE (pronounced "wee") is a system for automatic conversion of news articles on the Internet into video content presented in the style of TV news programs, using computer animation and speech synthesis (Fig. 3).

The relationship of text information in news articles to photos or other still pictures is determined (by finding the synchronized areas between the online graphics and text), and written expressions are converted to spoken expressions. In addition, an "impressions dic-

Fig.3 *Snapshot of news program-like animation generated by wEE*

tionary" indicating the relationship of words to an impressions scale (applying a database of news in newspaper articles over the past 12 years) is used to determine an impression of the news from the title and first paragraph; that is, whether the news is sad, reassuring, or so on. Based on weighted relationships within an evaluative index incorporating such impressions and on the extent to which content is understandable, favorable, endearing, and so on (based on research with test subjects), the article can be presented as a computer-generated announcer changes his or her tone of voice and facial expressions, with corresponding changes to the background music[17].

This system would render it easy to check news and other web information of interest even while the user is busy with household chores or other work.

5.2 u-Pav

u-Pav[18] is a system for passive reception of web content on a mobile phone. In contrast to wEE, which is designed for passive viewing of web content on a computer enhanced by computer animation and speech synthesis, u-Pav combines online graphics, text animation, and speech synthesis in a presentation format for mobile phones. To present information drawn from the web on a small screen and in a straightforward manner, groups of keywords are extracted from the web content and text animation is used to show the relationship between these groups. As with wEE, impressions of the content are extracted, and background colors are used to signal an impression of the page — blue for sad topics, for example, or red for lively ones.

5.3 Web2Talkshow

Web2Talkshow is a system for automatic conversion of web content to video content resembling a comic dialogue[19](Fig. 4). As with wEE, computer animation and speech synthesis is employed, but in this system, web content written as declarative sentences is automatically converted to dialogue, and this dialogue is further enhanced with humorous touches. Keywords extracted from the web content and a computer dictionary are used to find puns and related words, dialogue is generated to follow a predetermined script with traditional elements of comic dialogue (including punch lines), and character animation, virtual camera work, and other effects are added to create comedians represented by animated characters. In addition, for a more endearing effect, speech synthesis simulating a particular regional dialect (e.g., from the Kansai region) is used. This system holds potential in enabling viewing of even difficult news articles in a way that is interesting and enjoyable to older people and young children.

5.4 Interactive e-Hon

Interactive e-Hon is a system for converting text information in electronic content such as web pages and news articles on the Internet into animation, to help children better understand the given content[20](Fig. 5). Animation is created by finding the correspondence between text information written in declarative sentences to prepared components, including animated characters and actions. Further, this text information is converted to more easily understandable expressions in the form of a parent-child dialogue presented along with the animation. Explanations using animation and dialogue between parent and child characters will help even in the learning of pre-school children.

Fig.4 *Conversion of web content to video content in the style of comic dialogue*

Fig.5 *Conversion of web content to storybook-style animated content*

6 Conclusions

We have introduced the main results of research undertaken by the Interactive Communication Media and Contents Group on the fusion of communication and broadcast content — in short, a content fusion project. The fusion of Internet and TV content enables the efficient acquisition of information from a vast amount of content and presentation in a variety of formats. Content can now be used in value-added ways that render it easier to understand. Although limitations of space require that we summarize each set of research results and any related research, we encourage readers to refer to the individual papers for further details. Some results presented here have been selected for further development by NICT, and development for practical applications is underway, for corporate technology transfer, commercialization, and other purposes. Significant potential is also noted for development aimed at home information appliances and in real-world and ubiquitous content applications throughout the civic infrastructure.

Acknowledgments

This research is partly derived from the Informatics Research Center for the Development of a Knowledge Society Infrastructure, under the 21st Century Center for Excellence (COE) Program at Kyoto University (project leader: Katsumi Tanaka). In addition to the authors, research results introduced in this paper were made possible through the contributions of Atsushi Iizawa, Yasushi Ogawa, Tadashi Araki, Zoran Stejic, and Ai Kato of Software R&D Group, Ricoh Co., Ltd., Masaki Hayashi, Mamoru Doke, and Narichika Hamaguchi of NHK Science and Technical Research Laboratories, and Makoto Yokozawa, Toru Hamabe, and Hiromi Uwada of Nomura Research Institute (titles omitted). The authors also wish to thank Chikako Sugiyama, Asuka Miyaguchi, and Maki Taninaka for a variety of research support.

References

1 P. Lyman, H. R. Varian, K.Swearingen, P. Charles, N. Good, L. L. Jordan, and J. Pal, “How much information?”, 2003.
http://www.sims.berkeley.edu/research/projects/how-much-info-2003/

2 Ministry of Internal Affairs and Communications (formerly MPT) Advisory Council on Convergence of Communications and Broadcast towards 21st Century: Change of Information Environment and Convergence of Communications and Broadcast,
http://www.soumu.go.jp/joho_tsusin/policyreports/japanese/group/tsusin/kankyou/index.html

3 Broadcast Law, http://law.e-gov.go.jp/htmldata/S25/S25HO132.html

4 Law Concerning Broadcast on Telecommunications Service,
http://law.e-gov.go.jp/cgi-bin/idxoutrefdata.cgi?H_FILE_RECNO=01201&START_P=0&END_P=0

5 TV-Anytime Forum, http://www.tv-anytime.org/

6 EBU Tech 3295 The EBU Metadata Exchange Scheme,
http://www.ebu.ch/metadata/pmeta/v0100/html/P_META1.0/P_META3.html#anchor6
7 Common Metadata System J/Meta3.0,
http://www.fmmc.or.jp/fmmc-html/jmeta/053105_press_release.pdf
8 M. Henzinger, et al, "Query-Free News Search", WWW2003, May 2003.
9 Q. Ma and K. Tanaka, "Topic-Structure-Based Complementary Information Retrieval and Its Application", ACM TALIP, Vol.4, No.4, pp.475-503, 2005.
10 K. Sumiya, M. Munisamy, and K. Tanaka, "TV2Web: generating and browsing web with multiple LOD from video streams and their metadata", ICKS2004, pp.158-167, 2004.
11 S. Uchihashi, J. Foote, A.Girgensohn, and J.Boreczky, "Video Manga: generating semantically meaningful video summaries", ACM Multimedia 99, 1999.
12 H. Miyamori, Q. Ma, and K.Tanaka, "WA-TV: Webifying and Autmenting TV Programs for Nex-generation Storage TV", IPSJ TOD, Vol.47, No.SIG(TOD30), pp.71-80, 2006.
13 B. B. Bederson, J. Meyer, and L. Good, "Jazz: An Extensible Zoomable User Interface Graphics Toolkit in Java", UIST'00, pp.171-180, 2000.
14 G. W. Furnas, "Generalized Fisheye Views", CHI'86, pp.16-23, 1986.
15 T. Araki, H. Miyamori, M. Minakuchi, A.Kato, Z. Stejic, Y. Ogawa, and K. Tanaka, "Zooming Cross-Media: A Zooming Description Language Coding LOD Control and Media Transition", LNCS3588, 16th International Conference on Database and Expert Systems Applications (DEXA2005), pp.260-269, 2005.
16 T. Yamaguchi, I. Hosomi, and T. Miyashita, Webstage, "An active media enhanced world wide web browser", In International Conference on Human Factors in Computing Systems (CHI'97), pp.391-398, 1997.
17 T. Kumamoto, A. Nadamoto, and K. Tanaka, "Automatic Generation of TV News Program Explicitly Expressing Emotions", DEWS2006(Oral), 5C-oi3, 2006.
18 H. Uwada, A. Nadamoto, T. Kumamoto, T. Hamabe, M. Yokozawa, and K. Tanaka, "Automated Content Transformation with Adjustment for Visual Presentation Related to Terminal Types", Lecture Note in Computer Science 3841, The 8th Asia Pacific Web Conference (APWeb2006), pp.1214-1217, 2006.
19 A. Nadamoto, A. Jatowt, M. Hayashi, and K. Tanaka, "Web2Talkshow: Web content Transformed into Humorous Dialogue-based TV-program-like Content", Proceedings of the Intelligent Technologies for interactive entertainment (INTETAIN 2005), Springer, LNAI3814, pp.256-261, 2005.
20 K. Sumi, M. Nagata, and K. Tanaka, "Parent-Child Agents Information Presentation Model for Animation Media Transformation System "Interactive e-Hon", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.18, No.2(20060415), pp.240-250, 2006.
21 H. Miyamori, S. Nakamura, and K. Tanaka, "Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web", ACM Multimedia2005, pp.853-861, 2005.
22 H. Miyamori, Z. Stejic, T. Araki, M. Minakuchi, Q. Ma, and K. Tanaka, "Integrated Search Engine of Web Content and TV Programs", DEWS2006, 5C-i5, 2006.

***MIYAMORI Hisashi**, Dr. Eng*

Senior Researcher, Knowledge Clustered Group, Knowledge Creating Communication Research Center (former: Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Video Analysis, Natural Language Processing, Multimedia Information System

***KUMAMOTO Tadahiko**, Ph.D.*

Associate Professor, Faculty of Information and Computer Science, Chiba Institute of Technology (former: Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Spoken Language Dialogue system, Impression-Based Music-Retrieval System, Impression Mining From Text, CG Animation Generation

***NADAMOTO Akiyo**, Ph. D.*

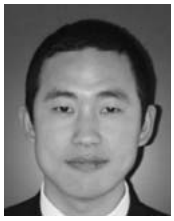
Senior Researcher, Language Grid Project, Natural Language Processing Group, Knowledge Creating Communication Research Center (former: Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Web Computing, Information Distribution and Browsing of Multimedia Content

***SUMI Kaoru**, Ph.D.*

Research Fellow, Universal City Group, Knowledge Creating Communication Research Center (former: Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

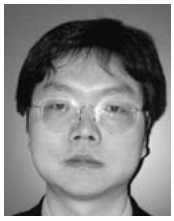
Human Computer Interaction, Knowledge Engineering



***NAKAMURA Satoshi**, Ph. D.*

Assistant Professor, Graduate School of Informatics, Kyoto University (former: Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Human Interface



***MA Qiang**, Ph. D.*

Assistant Manager, Ubiquitous Intelligence Technology Group, Service Platforms Research Laboratories, NEC (former: Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Information Retrieval, Multimedia Information System, Information Integration

***MINAKUCHI Mitsuru**, Dr. Eng.*

Research Fellow, Universal City Group, Knowledge Creating Communication Research Center (former: Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Interactive System, User Interface

***TANAKA Katsumi**, Dr. Eng.*

Professor , Graduate School of Informatics, Kyoto University (former: Group Leader, Interactive Communication Media and Contents Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Division)

Databases, Multimedia Information Systems, Web Information Retrieval