Aspects to keep in mind for the ideal Ubiquitous Web Information Retrieval Solution based on Agents

Andrea Paola Barraza  
Pontificia Universidad Javeriana  
Calle 40 #5-50.  
Edificio José Gabriel Maldonado, S.J.  
(571) 3208320 ext. 5392  
abaraza@javeriana.edu.co

Angela Carrillo Ramos  
Pontificia Universidad Javeriana  
Calle 40 #5-50.  
Edificio José Gabriel Maldonado, S.J.  
(571) 3208320 ext. 5392  
angela.carrillo@javeriana.edu.co

ABSTRACT
The rapid and uncontrolled growth of the World Wide Web (WWW) imposes severe constraints when meeting the diverse needs of today’s information society. The main challenge resides on: location, retrieval, integration and coherent presentation of quality and relevant ubiquitous information (anytime, anywhere and anyhow - with the correct format for the required device) from WWW, quickly and accurately, to satisfy a user’s singular information needs. Taking into consideration the unique features of the WWW environment, the limitations of existing solutions and the fact that agents are efficient tools for the development of Web Information Retrieval (WIR) systems; the current paper describes UWIRS (acronym of Ubiquitous Web Information Retrieval Solution), a web information retrieval system that includes aspects to keep in mind for the ideal WIR solution.

Categories and Subject Descriptors
H.3 [INFORMATION STORAGE AND RETRIEVAL]: Information search and retrieval – Information filtering, Retrieval models, Search process, Selection process.

Keywords
Multi-Agent System, Search Engine, Information Retrieval, Personalization, Adaptation, Filtering.

1. INTRODUCTION
At its early stages, the Internet was developed to satisfy information needs for the government and educational institutions, providing a secure small scale communications network. Because of its limited size scope and fast unexpected growth, the WWW has become a massive, chaotic, global, highly distributed center of information composed of heterogeneous (mostly incompatible) data sources. Aside of its obvious benefits (data distribution, communication, and scalability, amongst others); the WWW imposes severe constraints when meeting the diverse needs of today’s information society. When searching the web using traditional search engines (such as Google, AltaVista, Yahoo, Bing, and others), the researcher is usually answered with huge amounts of data, however this does not always mean that relevant, useful, quality information has been provided [19]. Consequently, researchers spend a considerable amount of time locating and integrating the data they actually need from the retrieved results, a dreary and wearisome task [8]. All the same, because the amount of information is overwhelming, desired information still remains unfound and hidden within unnecessary and irrelevant data [2].

It is clear then, that WWW information retrieval predicament demands new techniques to fulfill modern society’s information requirements. The main challenge resides on: location, retrieval, integration and coherent presentation of quality and relevant ubiquitous information (anytime, anywhere and anyhow - with the correct format through the required device-) from WWW, quickly and accurately, to satisfy a user’s singular information needs [19][6]. As a result, a search engine should answer a user’s query with the least amount of non-relevant web references as possible [12].

This paper pretends to lay out aspects to keep in mind for the ideal web information retrieval solution. The paper is organized as following: Section 2 illustrates a small overview, related to information retrieval, of the shortcomings of current solutions (specifically search engines) and the challenges presented by the WWW environment; Section 3 explains why agents are a well suited approach for a Web Information Retrieval (WIR) solution; Section 4 describes the functionality of the Ubiquitous Web Information Retrieval Solution (UWIRS), a WIR system that includes aspects needed for the ideal web search solution; Section 5 discusses related works; and finally, Section 6 presents conclusions and future work.

2. BACKGROUND
The WWW is relatively a new technology. Even so, in its somewhat short lifespan, it has experienced extreme uncontrolled growth, raising problems to a critical level [16]. The following section will illustrate only a small overview, related to information retrieval, of the shortcomings of current solutions (specifically search engines) and the challenges presented by the WWW environment.

2.1 World Wide Web Environment
According to Hoeber[12], even though the web could be seen as a single network with distributed repositories, it has features that make many of the traditional approaches to information retrieval, impractical and even impossible to put into practice. Some of these features are:

• **Huge and complex:** The size of the WWW is in the order of thousands of terabytes and grows at an extremely high rate. The lack of real control allows any individual to create and publish a web site with the information and structure, he or she chooses. In turn, the complexity related to each web site is higher than that of traditional text [10].
● Not secure and un-trustworthy: Because any entity can publish content on the web it is an un-secure network where the quality of the information provided is generally not guaranteed.

● Highly distributed heterogeneous data sources: Heterogeneous and distributed data sources contain diverse and mostly incompatible media with distinct formats [8]. The information gathered from these sources must be integrated and presented to the user in a coherent way.

● Network connections: Drawbacks of the wireless communications environment, protocol limitations, dissimilar bandwidths between users and service providers, multiple data formats, all impose restrictions over the use of the WWW network connections and communications [8] [17] [18]. Moreover, if the environment is mobile then it is hostile; it has low bandwidth, high rate of disconnection, high error rate, among others [18].

● Highly dynamic: Not only does it have a high growth rate; information is changed or updated almost constantly [10].

● Serves a wide variety of purposes for the different users: WWW users are very dissimilar, they have different backgrounds, interests, and come from different places; therefore have diverse information needs [10]. Additionally, individual users have access to different resources and place different priorities on different kinds of information (some users prefer to sacrifice response time to ensure quality information retrieval [11]).

● Only a small portion of data from the WWW is relevant or useful to a particular user: “It is said that 99% of the Web information is useless to 99% of the users” [10]. This may not be particularly obvious, but it is true that a certain user is only interested in a small fraction of the web.

Each of these features present challenges and considerations that have to be well thought-out for the development of WIR systems, such as: the decision of which small portion of the web is relevant to a particular user, the determination of strategies to identify high quality information, the provision of personalization and adaptation services for WIR, among many others that will be discussed in section 4.

In order to overcome the complexities of information retrieval from WWW, it is more than clear that tools must be developed to aid the user. Search engines are generally web sites that provide services related to the search and retrieval of data, from a certain portion of the web, given a number of keywords by the user. Because search engines are the most powerful and used WIR solution to date, they will be discussed in depth in section 2.2.

2.2 Search Engines

Search engines are and will continue to be, the primary tool used to find information on the web, due to the WWW’s complex characteristics (reference section 2.1). For instance, it has been reported that “88% of Web users start with a Web search engine when provided with a task to complete using the Web” [12] and that “85% of Web users find new Web pages using search engines” [14]. It is clear then; that web search engines have a dramatic effect on society. For example, in education, online articles have higher citation rate than does that are not online [13]. Search engines influence how information is used to a point that some consider these to be security and privacy risks. They also have several shortcomings associated to their search and information retrieval techniques, such as:

● Small search coverage: Search engines do cover massive amounts of information. Even so, compared to the size of the entire web, search coverage is still very small [19].

● Search precision is low: Most part of the results will have to be filtered by the user [19]. An example can be seen in Figure 1.

Figure 1. Example of low search precision in Google search engine.

● Lack of personalization: If different users (with unique information needs) type the same search query (using the same keywords) the search engine will return identical results. This is due to the search engine’s inability to analyze user behavior and determine search intentions or user preferences [19]. On the other hand, a standard search engine will treat a user the very same way as if he or she were using its services for the first time (there is no learning and therefore no personalization)[2].

● Generally don’t offer any ‘push-services’ or recommendation services: Without “push-services”, if the user is not active no information will be sent to him [2]. Recommendation and push services are useful to supply information needs users are not aware they have. Additionally, users turn to recommendation services when faced with a large array of choices such as does provided by search engines [4].

● Lack of Web Retrieval Support Systems: Search engines are established under the assumption that users are able to accurately portray their information needs and correctly browse and filter given results. Nonetheless, if a user is not familiar with domain-specific keywords he or she won’t be able to craft his or her query, a precise statement of the users information desires won’t be delivered, and therefore, the search won’t be successful[16]. Researchers need support in activities such as refining queries (query enrichment is usually a task of the researcher), exploring results, filtering, organizing retrieved documents, amongst other activities performed with the retrieved information. In summary, search engines lack Web Retrieval Support Systems that focus on giving support to user-centric tasks performed along with the web search [12].

● Lack of adaptation: Web information retrieval services are empowered when the application has the intelligence to process contextual information (location, type of connection, adaptation to the characteristics of the used device –WIR systems generally don’t adapt to mobile devices -, and others) [6].

● Need of web tour guides: Once a user has selected a link from the given results after using a search engine, they step out of the search engine’s boundaries and are left to continue the search on their own [16]. Search engines don’t give ongoing navigational support and in most cases only a sequence of pages can satisfy the user’s information needs [16]. Systems such as WebWatcher [15] and SWAMI [16], intend to provide solutions that accompany the user from page to page, make suggestions of appropriate hyperlinks, and learn from experience to improve their services [16] [15] (for more information view section 5). They guide the user along the appropriate path much like a tour guide conducts a tourist through a museum [15].
It is important to highlight these flaws, since they will help pinpoint areas that require improvements over current solutions, or will stand as considerations to be taken into account when creating innovative solutions that move towards the next-generation of web search systems.

The need for new solutions that support WIR is clear, now it is relevant to mention appropriate tools for the construction of these solutions. The current paper focuses on an agent based system; for this reason, section 3 only explains why agents are well-suited tools for the development of WIR solutions.

3. WHY AGENTS FOR INFORMATION RETRIEVAL?

Multi-Agent systems enhance the efficiency of intelligent search and are especially well-suited for distributed environments such as the WWW [17]. An agent possesses the following desirable qualities for WIR solutions:

- **Autonomous**: Agents can be independent and make autonomous decisions on a non-deterministic way; allowing the parallel execution of multi-tasks in a distributed environment like the WWW. This faculty allows the delegation of tasks (reducing the complexities of the application domain) where minimal supervision/intervention is needed for the agent to carry out its activities [19][14][2][17].
- **Pro-active, reactive and goal-oriented**: Agents must be able to reason about their environment (network changes, user activities, among others), learn and make decisions towards a set of goals in a proactive (by its own) or reactive way (in response to an event) [2][6].
- **Adaptive**: The WWW environment is highly dynamic, and an agent residing on the web must be able to adapt or adjust to its context so it can correctly carry out its goals and objectives [6].
- **Mobile**: Agents may choose to migrate towards another location to accomplish certain tasks; thereby reducing network load [19][14]. Special considerations have to be taken into account for mobile agents (e.g. security). It’s important to clarify that not all agents on the web are mobile agents.
- **Perseverant**: Agents may take on tasks covering a long period of time [14].
- **Sociable**: Agents may communicate with other agents through a predefined agent communication language, while a user interface agent can interact with a web user in a language that is much more natural for the user. Communication allows agents to cooperate with each other or with a web user, and undertake certain tasks [14].

Significant amount of literature and research on Web agents ensure that these are better tools for WIR development [13]. This paper is based on the premise that Multi-Agent systems are one of the best approaches for WIR solutions, and the following section will describe the ideal web search system based on a Multi-Agent approach.

4. IDEAL WEB SEARCH

Taking into consideration the unique features of the WWW environment, the limitations of existing solutions and the fact that agents are efficient tools for the development of Web Information Retrieval (WIR) systems; the current paper describes **UWIRS** (Ubiquitous Web Information Retrieval Solution), a web information retrieval system that takes in to account aspects to keep in mind for the ideal WIR solution.

To accomplish the set out task, the problem of web information retrieval was divided into three stages: (1) Before initiating the search process (understanding by search process the actual sending of the interpreted user query towards the correct information sources) (2) During the search process (3) After information retrieval. Each of these stages will be explained on sections 4.1, 4.2 and 4.3 respectively. It is important to point out that the described system assumes the environment provides the necessary support for the creation of agents, communication
between agents, white and yellow pages services administered by a Platform Agent, agent migration, protection and destruction of agents [18].

The UWIRS’s logical agent architecture, where the different UWIR’s agents and their interrelationships are portrayed, can be seen in Figure 2. It can also be seen that the selected agent organization clearly separates three different environments: the client-side, the UWIRS platform, and the information providers’ side. The following sections will explain the different agents residing on each environment, giving special consideration on the subjects of: agent interactions, knowledge of an agent, protocols (agent to agent, agent to user, agent to information provider), goals, and learning capabilities [3].

For the purposes of this article, the UWIRS’s scope will concentrate solely on the task of web information retrieval. Services such as recommendations, push services, pre-catching, cache managing and web tour guides, will initially not be considered. Nonetheless, UWIRS could be extended to include these services in future works.

4.1 Before initiating the Search Process

In UWIRS, when a user needs to carry out a particular search request he initially has one of two choices: sign in as a recurrent user or as a visitor. When the user logs in as a recurrent user, the system provides services that are adapted to a Device-Profile (containing information of the used device) and the user’s User-Profile (containing the system’s knowledge of the user). When the user logs in as a visitor, the system must devise ways to create an initial User-Profile as a foundation for future services. In UWIRS, the initial User-Profile is assigned based on the results given from a user-feedback form and is only available for the visitor’s current session. A session starts when the user initiates interaction with the system (signs-in) and ends when the user decides to end the session. If the user does not interact with the system for an extended period of time the session will time-out and also end.

In an ideal system, a communication protocol must be worked out to facilitate the user’s interactions with the system (specifically with an agent), allowing a user-friendly intuitive data request interface. It is important to provide means for the user to specify his/her information needs in his/her preferred mode of communication (user natural language request by: speech, text, graphic, among others) creating a multi-modal interface [14]. The user could also provide examples as the starting point for a web search [2] (a paper, a picture, a webpage, among others) for which the results would include related or similar documents. Additionally, a support system must assist the user in the specification of the search query, this way it is ensured that the user is a minimum source of errors. Finally, the user’s search request must be sent to a UWIRS platform (through the Client-Session Agent), where it is processed and translated into a UWIRS query language and used to initiate a search process. When the search process is finished a set of results is returned to the user and displayed, by the User-Interface Agent, in the user’s preferred mode of communication (as seen in Figure 3).

To manage any kind of user-interactions, a specialized agent residing on the user’s entry device (usually attached to a browser [16]) is created (one for every user connected to the device), called User-Interface Agent. This agent must cooperate with the user’s User-Profile Agent and with the associated Device-Profile Agent to determine a way to provide custom-made services adapted to the user and his/her used device. In addition, the User-Interface Agent must also report his observations of the user’s interactions with the system, to a User-Profile Agent residing on a UWIRS platform. This way an upgraded User-Profile is maintained by the system.

At this point the UWIRS system must have a way to represent the knowledge managed by each agent (including the profile set: User-Profile, Device-Profile, Information-Provider Profile) such as Ontologies represented by XML files managed by JESS databases. Moreover, a communication protocol must be established to allow the interaction between the agents residing on the client-side (used device) and on the UWIRS platform (an agent to agent communication protocol).

In order to manage communication for the current session between a UWIRS platform and the client, an agent must be created in the client-side as well as in the server-side. In UWIRS these agents will be called Client-Session Agent and Platform-UserSession Agent respectively. Both agents will cooperate to manage the user’s session properties, user’s device properties, and the connection properties (type of connection, restrictions, type of network, communication protocols, among others) [6]. This way transparent communication can be carried out between a point-of-entry device (desktop computer, cell phone, personal digital assistant, or other compatible device with internet access) and a UWIRS platform (see Figure 3).

An additional responsibility of the Client-Session Agent is to offer information to the associated Device-Profile Agent (residing on a UWIRS platform) about the device used to access the UWIRS system (generally in the form of Composite Capabilities/Preferences Profile – CC/PP – documents). This way an upgraded Device-Profile is maintained by the system.

On the other side, it is important to note that the Platform-UserSession Agent’s main responsibility is to represent the user in the UWIRS system [6]. Because the Platform-UserSession Agent can work on behalf of the user, the system has support for disconnected operation and their by can reduce network load [9].

Before moving towards the description of Stage (2) During the search process, the use of User-Profiles, Query Processing and the management of Information Source Profiles must be explained in

1 http://www.w3.org/Mobile/CCPP/
depth. These concepts are described in sections 4.1.1, 4.1.2 and 4.1.3 respectively.

4.1.1 User-Profile
The biggest challenge for a WIR solution is to present the user with useful, quality, short and precise information the user is really interested in, anytime, anywhere, anyhow. To accomplish this, the system must learn and model the user's interests (long term, short term or periodic [16]), preferences, needs, behaviors, contextual information (e.g., user location), history, interactions with the system and other attributes that make the user unique; in a User-Profile. A UWIRS's User-Profile is composed of two types of information: static (e.g. date of birth) and dynamic (e.g. user interest). Using this profile, the system can provide a superior user experience with services tailored to the specific user, such as: user-centric web searches [16], possible push services, suggestions about possibly interesting web pages, and others.

In UWIRS, a specialized agent is assigned the task of managing, updating and controlling the User-Profile; called User-Profile Agent (see Figure 3). The User-Profile Agent is informed by the User-Interface Agent of the user's interactions with the system. By correctly interpreting the received reports of user actions, the profile will be updated to contain information learned from implicit (e.g., user interactions with the system) or explicit (e.g. user input such as rankings) user feedback[2].

A User-Profile may be associated with one or many User-Roles (as seen in Figure 3). Examples of User-Roles could be: Student, Teacher, Employee, Father, Mother, Computer Programmer, among many others. User-Roles could be viewed as the representations of generalized stereotyped User-Roles and the User-Profile contains customizations of these stereotypes [14]. During a user session one or many User-Roles are dynamically chosen to determine search properties (query-processing and filtering). The user may explicitly choose one or many of his User-Roles with whom he or she identifies best with or the system may infer and establish which User-Roles to use according to a deduced user context (Social Context, Personal Context, Task Context, Spatio-Temporal Context [17], others); at any point of the current user session.

4.1.2 Query Processing
The UWIRS platform holds a community of Query-Processing Agents. Query-Processing Agents are in charge of managing the whole query process: receiving user-query request through the Platform-UserSession Agent, interpreting and translating the user-request, enrichment of the interpreted query, correctly delegating information retrieval tasks, integrating information-retrieval results, correctly delegating filtering services of the retrieved results and finally sending the search results to the user through the Platform-UserSession Agent.

There are different types of Query-Processing Agents; each specialized depending on the received user-query format (e.g. voice, text, graphic). Each agent is in charge of the translation and interpretation of the user's search request to build a UWIRS query. After obtaining a UWIRS query, the Query-Processing Agent adds adaptation criteria [7] and enriches the query by cooperating with the User-Profile Agent and Device-Profile Agent (as seen in Figure 2).

After the enrichment process, the Query-Processing Agent decides if the current query is better solved by its division into sub-queries (depending on the queries complexity - Can it be solved by one or many information providers?) and proceeds to formulate optimized retrieval instructions (or a retrieval plan) for each of the possible queries. These retrieval instructions are built upon UWIRS knowledge of the different information sources in the WWW, provided by an Information-Provider-Profile-Manager Agent (reference section 4.1.3). For each needed query, the associated retrieval instructions are assigned to an Information-Retrieval Agent.

The Information-Retrieval Agent returns the obtained query results; and it is the Query-Processing Agent's task to integrate the returned results from the different Information-Retrieval Agents (in the case of user query division).

The Query-Processing Agent passes the combined results to different Filter Agents to improve and further adapt the search results (see Figure 2). The Filter Agents are chosen according to the User-Profile and selected User-Roles. Filter Agents return filtered results to the Query-Processing Agent who coherently joins all results. These, are sent to the client-side through the Platform-UserSession Agent who communicates with the Client-Session Agent who, in turn, communicates with User-Interface Agent who finally displays the results to the user.

4.1.3 Information Source Profile
Before continuing, it is important to explain how UWIRS obtains and maintains upgraded Information-Provider Profiles and the information these enclose.

An Information-Provider Profile contains information associated to the Information Provider, such as: content description (metadata), important relationships with other information providers, facts related to issues of trust and reputation [6], and associated Wrapper-Agents. Wrapper-Agents help translate the information provider’s interface to one a UWIRS’s agent can understand.

There are three types of agents involved in the process of building and maintaining an Information-Provider Profile: Information-Provider-Profile-Manager Agent, Web-Search Agents and Information-Provider-Observer Agent (see Figure 4). The Information-Provider-Profile-Manager Agent, is the Information-Provider Profile administrator, and is in charge of correctly updating and adding information provided by the Web-Search Agents or Information-Provider-Observer Agents. This agent also arranges the Information-Provider Profiles in a strategic way that facilitates searching between them (e.g. creates clusters).

On the other hand, Web-Search Agents are continuously searching the web (a search strategy must be defined for the Web-Search Agents) for new information providers, with the help of the Wrapper-Agents’ services. Once a Web-Search Agent detects a new information provider it informs an Information-Provider-Profile-Manager Agent who creates a new Information-Provider Profile with information given by the Web-Search Agent. If possible, the Web-Search Agent creates an Information-Provider-Observer Agent on the information source. The Information-Provider-Observer Agent is in charge of informing changes on the information provider to the Information-Provider-Profile-Manager Agent so an updated Information-Provider Profile is kept.

The main challenge when building Information-Provider Profiles is to extract information without the need of special tagging or services on the information provider [5].
4.2 During the Search Process
The search process is the actual sending of the UWIRS query towards the correct information sources that have been defined by the retrieval instructions developed by the Query-Processing Agent. The agent in charge of retrieving information and dynamically following the retrieval instructions is the Information-Retrieval Agent. The sentence “dynamically following the retrieval instructions” is used because the Information-Retrieval Agent may change the retrieval plan as it is followed, deciding upon their next action depending on the current context and retrieval factors (e.g. network speed, data size, resources provided by data servers [8]). Information-Retrieval Agents may also decide to communicate and collaborate with other Information-Retrieval Agents to obtain common goals.

As a way of establishing a communication-protocol between an Information Source and an Information-Retrieval Agent a mediator is needed. An Information-Retrieval Agent may request the services of a Wrapper Agent that will help translate the Information Provider’s interface to one the Information-Retrieval Agent can understand. This greatly simplifies the development of the Information-Retrieval Agent because they only need to know one communication protocol, making it possible to scale the network and have many Information-Retrieval Agents assigned to different types of information sources [3].

After Information-Retrieval Agent completes its assigned retrieval instructions (retrieves the information associated to the assigned query), it returns the results to the Query-Processing Agent so it can continue the query process (see Figure 4).

4.3 After Information Retrieval
This stage begins when all Information-Retrieval Agents have returned their query results to the Query-Processing Agent, who decides to start a filtering process. UWIRS has a community of expert Filter Agents (each agent is a specialized expert on a certain kind of filtering technique) for this task (as seen in Figure 5). A Query-Processing Agent selects the correct Filter Agents (e.g. Display Filter Agents, Content Filter Agents) for a certain query process, depending on the current User-Profile (along with the selected User-Roles) and on the associated Device-Profile. If the selected Filtering Agents are not able to sufficiently filter all the retrieved search results, support can be found on collaborative filtering, for example: if a web document was rated interesting by a user with a similar User-Profile then it might be of interest for the current user [2].

5. RELATED WORKS
This section describes four projects that are similar to UWIRS, their advantages and disadvantages.

SAIRE [14] (Scalable Agent-based Information Retrieval Engine) is a Multi-Agent search engine that provides non-expert users with transparent access to Earth and Space Science data, maintained by NASA2 and NOAA3 in distributed and heterogeneous data sources, over the internet. SAIRe provides natural-language understanding and an integrated user interface that hides complex information retrieval protocols from users. Because SAIRe is limited to NASA and NOAA data sources, it’s able to provide domain-specific protocols, used by Domain Specialist agents, who therefore have conceptual information retrieval capabilities. SAIRe also provides dynamic adaptation to user preferences which helps the system reason about the user’s objectives and intentions. Initially the user is modeled by means of stereotypes and his individual profile is then customized with the individual user’s particular information. SAIRe’s Multi-Agent Architecture holds a centralized control scheme where a Coordinator Agent controls the activities of the entire system. Even though this provides some advantages for the system, it is clear that the Coordinator Agent is a single point-of-failure. Other disadvantages of SAIRe to point out are: its limited scope (does not cover the entire web); its lack of query division techniques; its lack of support for information filtering; and the inability to adapt retrieved information to the user’s chosen entry device.

ACQUIRE [8](Agent-based Complex QUerying and Information Retrieval Engine) acts as an interface agent that allows a user to pose high-level declarative queries to large, heterogeneous, and distributed data sources as if they were a single, unified, homogeneous data source. Each user query is translated into a set of sub-queries, using site and domain models of the distributed data stores (tests were initially done with three NASA data archives). ACQUIRE spawns a set of mobile agents with an assigned retrieval plan for each sub-query. Each agent has the option to dynamically change the optimized retrieval plan depending on retrieval factors (e.g. network speed and data size); and can communicate and collaborate with other agents to decide where to send data or migrate. Finally when all mobile agents

---

2 NASA- National Aeronautics and Space Administration
3 NOAA- National Oceanic and Atmospheric Administration
return, ACQUIRE merges and filters the retrieved data and presents the results to the user. Among ACQUIRE’s advantages, the use of sub-queries, data and site modeling and dynamic retrieval plans can be highlighted. Disadvantages to point out include: the use of high-level declarative queries instead of Natural Language Processing (NLP), its limited scope (does not cover the entire web), and the lack of adaptation (to the user and the used device) techniques.

PIA [2] (Personal Information Agent) is an agent-based personal information system that collects, filters and merges data to deliver the right information at the correct time, offering access to information through the WWW, e-mail, SMS, MMS and J2ME clients in a situation-aware manner. Combining push and pull techniques, PIA allows a user to: (a) explicitly search for specific information, and (b) be informed automatically about relevant information divided into groups (pre-, work, and recreation). A personal agent manages a user’s individual profile, offering adapted information and learning from user feedback. On the other hand, Information-extracting agents constantly gather new information from a variety of sources (internet, databases, web-services among others). Additionally, the PIA system uses advanced filtering techniques organized in multiple filtering communities that allow collaborative and content-based filtering. One of PIA’s biggest setbacks is that it only searches for information in text format and does not adapt to the user’s preferred entry device. On the other hand, PIA does not support query division, query enrichment or the development of a query retrieval plan.

The Intelligent Information Retrieval Model [19] (IIRM) was designed with a Multi-Agent approach to assist users in finding interesting information in the WWW. IIRM searches through popular web search engines, filters the retrieved information, and provides a user with a small number of results that are most likely to be relevant to the particular user. The filtering process is carried out by an agent that analyzes the user’s behavior and infers search intents and preferences. A flaw to IIRM’s approach is its dependency on the quality of service of other systems (other search engines). On the other hand, generally search engines look in the same small portion of the web, so the difference between results provided by different search engines is not very likely to vary, which means IIRM still has a limited scope. Even so, IIRM must still answer: How many search engines to be used are enough to provide complete results?

Table 1. Related Works comparison to UWIRS.

<table>
<thead>
<tr>
<th>MOBILE AGENTS</th>
<th>SAIRE</th>
<th>PIA</th>
<th>ACQUIRE</th>
<th>IIRM</th>
<th>UWIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTENT ADAPTATION</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PERSONALIZATION</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QUERY ENRICHMENT</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FILTERING TECHNIQUES</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>NLP</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>MULTIMODAL INPUT</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>SUB-QUERIES</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>RETRIEVAL PLAN ROUTING</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DOMAIN MODELING</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In Table 1, a clear comparison between attributes and functionalities of the presented related works and UWIRS can be viewed. The symbol ‘+’ indicates the criteria is included, the symbol ‘-’ indicates not included, the symbol ‘?’ indicates that it isn’t clear if the criteria is included. It can be concluded from this table (Table 1), that even though each of the related works meets some or most of the criteria, none are totally satisfactory as an ideal WIR solution. Only UWIRS considers all attributes.

6. CONCLUSIONS AND FUTURE WORK

The WWW can be viewed as a colossal library where documents are not organized and therefore desired information is difficult to find. In one of his short stories, “The Babel Library”, the writer Jorge Luis Borges describes infinities of libraries containing books with unintelligible titles such as “Axaxaxas milo”. People walked aimlessly through the halls until dying, building hypothesis about the things they could have found but didn’t. This is a cruel analogy of the current problematic to be confronted towards the World Wide Web: we live in a growing world of data where lots of facts exist and little information can be found in [1]. Today’s information society demands new and innovative solutions not just for information retrieval but for knowledge retrieval.

This paper has presented aspects to keep in mind when developing a web information retrieval system by describing an agent based solution called UWIRS. For UWIRS to become a possible reality, some problems in the WWW have to be solved, like: the creation of standards for mobile agents (to solve problems such as security), privacy issues related to the use of information, solutions for the associated restrictions viewed when trying to translate and understand an information provider’s content, among others. Possible solutions include the establishment of control organizations that can provide and enforce the use of standards. Even so, information retrieval systems must try to mitigate these problems.

Finally, it is hoped that by establishing the main features and “must-haves” of an ideal web information retrieval system, UWIRS serves as a stepping stone for the creation of new generation WIR solutions. For future tasks, it is important to add to UWIRS services that weren’t included, such as: recommendations, push services, pre-catching, cache managing, and web tour guides. On the other hand each of the mentioned aspects is an area for future investigations, from which the following stand out: query enrichment, content filtering, information acquisition for the profile set (User-Profile, Device-Profile, Information-Provider Profile), and communication protocols (agent to agent, agent to user, agent to information provider).

UWIRS will be implemented using Java and the agent-based framework BESA [9] (acronym of Behavior-oriented, Event-driven, and Social-based Agent). BESA provides a series of favorable attributes for the construction of social-based agents such as the agents found in UWIRS. Furthermore, the BESA framework is developed upon an internal agent architecture that offers support of concurrent agent behaviors which respond to events, allowing agents to interact with other entities (other agents or with other objects in their environments) using an event-driven communication system. We will validate UWIRS by means of a prototype intended for an educational environment. The prototype will be considered successful, if a specific group of students using UWIRS are able to retrieve data that meets their information and
learning needs, from different information sources, taking into account contextual features.

REFERENCES


