The Web of People: A dual view on the WWW

Michel PLU

France Telecom R&D
2 avenue Pierre Marzin
F-22307 Lannion CEDEX, France
+33 2 96 05 36 98

Michel.plu@rd.francetelecom.com

Pascal BELLEC

France Telecom R&D
2 avenue Pierre Marzin
F-22307 Lannion CEDEX, France
+33 2 96 05 31 42

Pascal.Bellec@rd.francetelecom.com

Layda AGOSTO

France Telecom R&D
2 avenue Pierre Marzin
F-22307 Lannion CEDEX, France
+33 2 96 05 36 98

Layda.Agosto@rd.francetelecom.com

Walter VAN DE VELDE

CampoRosso
67 avenue Ducpetiaux
B-1060 Brussels, Belgium
+32 2 538 81 15

wvdv@camporosso.be

ABSTRACT

In this paper, we describe an alternative web paradigm, called the Web of People. Rather than a web of pages, it is a web of persons. The paradigm is motivated by a number of obvious shortcomings of the current Web, in particular its tendency to provide centralized support tools for indexation and search. Instead our model exploits the distributed intelligence of the users. We describe the basic principles of the Web of People, its architecture, and its technical realization in a prototype system called SoMeOne. In particular, we elaborate on a search engine for the Web of People, aimed at finding people that act as relevant information sources.

Categories and Subject Descriptors

H4.3 [**Communications applications**]:

General Terms

Algorithms, Management, Experimentation

Keywords

Communities' support, personal indexing technologies, information sharing, resources discovery, social media

1. NEW WEB PARADIGMS

In his book 'Internet Dreams' [1] Mark Stefik explored in depth four internet metaphors: the Library metaphor, the Mail metaphor, the Marketplace metaphor and the Worlds metaphor. Each such metaphor corresponds to an archetype of a human ideal: the keeper of knowledge, the messenger, the trader and the explorer. Exploring metaphors, Stefik argued, is useful since they provide powerfully coherent ways to think about the internet, implying also that the way one thinks about it will influence what it will become. Moreover, the implied archetypes put the question of authenticity: do these metaphors reveal what we really want to be?

The Web of People, which we present in this paper, is one element in a wider study in which a dozen future Web paradigms are explored, much in the spirit of Stefik's work. Our motivation has been to understand what the WWW may become, rather than understanding how current problems with it may be addressed.

The Web of People, as we will see, provides a *dual* view of the current Web. Rather than humans navigating from information to information, it is the information that navigates from human to human. Our key idea is to exploit the distributed intelligence of users and their network of acquaintances as a resource for creating a meta-view of the Web which, we think, will be much more adapted to individual needs than what is the case with current centralized search and indexing engines.

In the Web of People the role of each and every user as an autonomous, competent and responsible mediator of information is recognized. It capitalizes on the human need to be perceived and recognized as part of a network of information exchange, be it for knowledge, news, hearsay or rumor.

1.1 What is a Web ?

How to characterize the larger class of systems of which the current World-Wide Web is a prime instance? Answering this requires setting somewhat arbitrary boundaries, but the following is a convenient starting point:

> *A Web is a system of nodes and relations, providing content and a structure that, by some process of interaction, can be interpreted as meaningful (and thus somehow useful) information.*

Moreover, at least the following features seem essential for the current WWW experience, and are preferably preserved:

It is <u>open</u> (anybody can contribute easily to its content and structure).

It is <u>heterogeneous</u> (variety of content types).

It is <u>ubiquitous</u> (available anytime, anywhere, anyhow)

It is <u>distributed</u> (non-centralized ad-hoc implementation and management).

It seems to be an unavoidable consequence of these features that a Web, when widely deployed, will be <u>non-static</u>, <u>growing</u>, <u>rich in content</u> and <u>complex in structure</u>. And exactly because of this the Web risks to be a victim of its own success. The Web is becoming

so overwhelming that, instead of satisfying the user, it can easily become a source of frustration and waste of time.

1.2 The current WWW

A Web paradigm provides a high-level conceptual framework within which the structure, behavior and usage of a future web can be understood and communicated. For instance, the 'original' World-Wide Web can be described as a large and distributed collection of pages of natural information that is navigated by users tracing hyperlinks and using search engines.

There are at least two characteristics of the current WWW that are not in line with its proclaimed 'spirit'. First, the WWW model is essentially a single user model. Although it is based on an infrastructure that can support millions of parallel users, they are basically operating in their own individual threads. The WWW is a given, and each user explores it independently of the others. This reality is in sharp contrast with the discourse of the Web as a social facilitator.

The second characteristic has to do with the organization of the Web. The Web is often heralded as being self-organized. Yet, it *is* being organized, be it in a largely distributed fashion. Most worrying, in a way, a handful of search engines and portals are dominating the use of the Web so that, whatever organization is being imposed by designers of web sites, it is rapidly surpassed by the meta-level view of your favorite search engine, portal or information directory.

1.3 Users as Peer Information Mediators

A wide variety of tools are being deployed or researched to make life on the Web easier. However, they are generally thought for creating centralized intelligent systems. Moreover, the needs to be addressed are far beyond nowadays' automatic analysis and indexing technologies. For example, how to identify, automatically, a document containing false information? How can software recognize that a level of document description is appropriate to a user's background knowledge, or to measure the clarity of the discourse, or the pedagogical qualities of its presentation? How can software model the user's sensibilities in order to detect funny stories, beautiful pictures, dramatic movies that he/she will be sensitive to and will certainly appreciate? The Semantic Web research community is addressing these issues, attempting to provide technological solutions that can cope with the quantity and diversity of information and users.

Our approach, however, is different. We use the distributed intelligence of the users that handle the information. We help users to exploit their relationship- and information exchange networks to find and filter information between each other. By doing this, we develop a kind of new network where information navigates from users to users instead of having users navigate through information; this is why we call this network the "Web of People". As with push technologies, information goes directly to the user. Nevertheless, instead of having channels controlled by information providers, we manage networks of human channels. With our system, a user from trustworthy relationships becomes a Third Party Confidence. We think that the selection made by these human channels can be much more personalized and adapted to users needs.

Our goal is not to replace traditional search engines but it is to complement them for example with new ranking mechanisms. Second we hope to scale with the number of the users that participate and that post new information in the network.

1.4 Overview

Section 2 describes the basic principles and technical architecture of a Web of People as we have implemented in a prototype system called SoMeOne. Section 3 describes in detail the concept and realization of a search engine for the Web of People. Section 4 provides early results from experimental use, while Section 5 discusses related work. The paper closes with a perspective on future work and conclusions.

2. THE WEB OF PEOPLE

2.1 Basic Principles

The basic principles of the Web of People are the following.

The nodes of the Web of People are the users. Each user stores and manages references to resources of the WWW. References are defined by URLs [2]. Those references connect the Web of People and the WWW. Each reference is classified according to a personal and hierarchical taxonomy. We call each element of the taxonomy a topic. A review is a reference associated to one or several topics with eventually a textual comment.

What are the relationships between the nodes in the Web of People? Each user can decide to send references in one topic to another user. For this purpose each topic in a user's personal taxonomy has a distribution list that contains a list of user references, thus creating as many information pipes between the corresponding nodes. The contacts of a user are all users sending information to that user and all users receiving information from that user. They form the direct neighborhood of the user node in the Web of People.

Instead of having to manage as many distribution lists as there are topics in her taxonomy, a user can reuse and specialize the distribution lists. A topic that is a specialization of a more general one by default inherits the distribution list of its ancestor. A user can always specialize this default list by adding or removing contacts. She can also declare that a topic is "secret". Any review associated with such a topic is only accessible to its author.

Whenever a user wants to better characterize a referenced resource on the Web, she adds a topic to the corresponding review. The Web of People then distributes the review through the Web of People, based on the distribution list associated with the topic. We refer to this distribution process as "semantic addressing", to contrast it with having to precise physical mailing addresses. Note that, according to the topics put together in a review, the distribution list will not be the same (for example if one topic is declared "secret" then the distribution list will be empty). Therefore, this semantic addressing process is also contextual.

A user accesses the Web of People through her personal node. This node provides her with her personal information and the information she is receiving. It also provides the facilities to manage the personal taxonomy of topics and the associated distribution lists.

Figure 1 shows the personal page of a user as it is displayed when he accesses his node. It shows in the left column the topics managed by the user. The middle column contains the received topics, which are topics the user can access and the number of reviews about new references in those topics. The right column displays the reviews in the selected "received topics" which, in this particular screenshot, is entitled "conf social media" (i.e.,

conferences about social media). The button at the bottom of the middle column ('Supprimer ce Topique') lets the user filter a received topic. Buttons at the bottom of the right column let the user categorize the selected resource with is own topics, indicate that he is not interested by the resource ('Delete' button) or that

he has seen the resource and does not want any new reviews about it ('Seen' button).

To prevent spamming a user does not directly receive information when she is added on the distribution list of another user's topic.
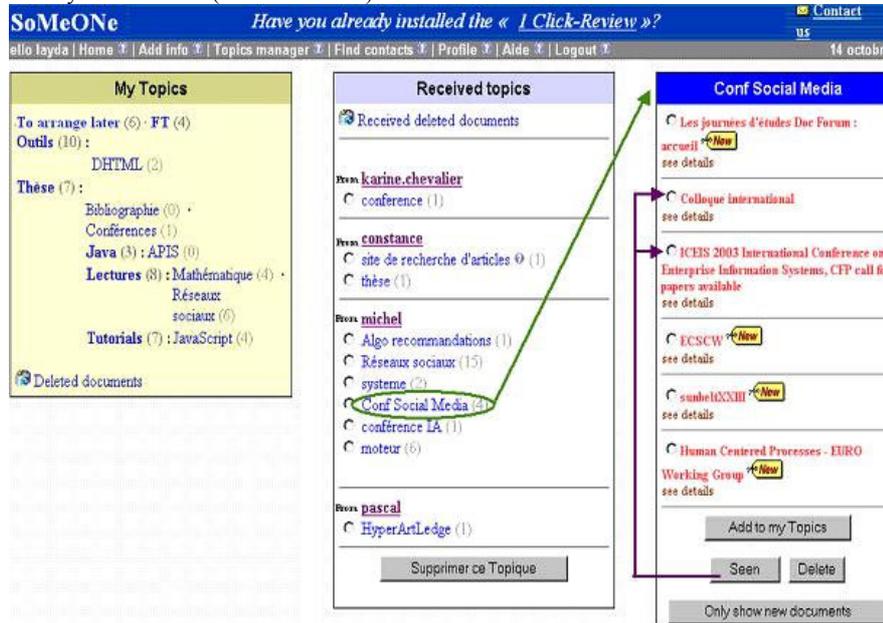

Figure 1: SoMeONe's home page.

She is only informed that she can receive information from a user about that topic. It is up to her then to decide to accept to receive the information or not.

Whenever a user receives a review, she has to decide whether to accept the information or not. To do so, she can consult the current review containing the topic. Moreover she can consult the information about the sender, which depends on a public profile managed by each user in the Web of People. At any time she can change her mind and refuse to receive any more information about that topic or from this user.

When a user receives a review, she may decide to store it in her personal node. Since each user applies an individual taxonomy of topics, this is done by adding some of her own topics to the review. As before, the review then becomes available to all the users who are in this distribution list of these new topics.

The Web of People is thus a peer-to-peer network of topical information push channels, where each channel is co-managed by both sides. Note that it does not require a globally agreed upon taxonomy of topics: every user creates her own view on the Web, by own research or by re-interpreting reviews from others in her own topic ontology.

2.2 Architecture of the Web of People

2.2.1 Personal server on a personal node

Each user has its own active personal 'Web of People' site. The site can be a traditional WWW site with an http server and a specific application server. This application server manages the

user information, controls the access to the user information and provides the user with the functionalities to manage his public profiles, personal taxonomy and associated distribution list. With this architecture, each server is always connected. Therefore, each piece of information is reachable at any time. Furthermore, users can access their information from any machine connected to Internet and any terminal like a PDA or through WAP.

In order to simulate and experiment the Web of People we used the Jalios Content Management Suit to develop the application server [3]. Of course if we want the Web of People to develop as the WWW develops we need to define lighter and more standardized solutions.

A personal server manages an XML file to store the user's information, reviews and topics.

A review is an entry in the XML data file. It contains the URL and meta-data about this URL: the title of the document, a comment and a list of associated topics.

A server must also contain a list of the users that gave at least one access to their topics. This list is called the "list of the contacts". It contains a list of user's identifiers (i.e., the URL of their personal server).

Each topic has an associated list that contains the identifiers of the users that can access its contents.

In order to share information between personal servers it is important that they are synchronized on the same date. The

solution we adopted is to use the Network Time Protocol [4] to synchronize the clock of each personal server.

2.2.2 Central user identifier server

In order to manage unique user identifiers and the associated addresses of their personal server we installed a central user identifier server (UIS for short). In order to scale up, a more distributed architectures of this name server has to be defined following DNS principles.

This UIS contains an XML data file that stores entries by user. Each entry contains the name of the user and the URL of his personal server (this URL serves as the unique identifier of the user). With this central server, then, it is possible to easily find the personal server of another user.

Each time a user installs a personal server on his web page, he needs to declare his name and the URL of his server to the central server. After that he exists in the Web of People.

2.2.3 Communication between a personal server and the central server

To give an access for a topic, it is necessary to display the list of users that are registered. A portlet displays this list. By making a comparison between the distribution list of a topic and the list of registered users, we can easily construct a page that shows who are the users that can access a given topic.
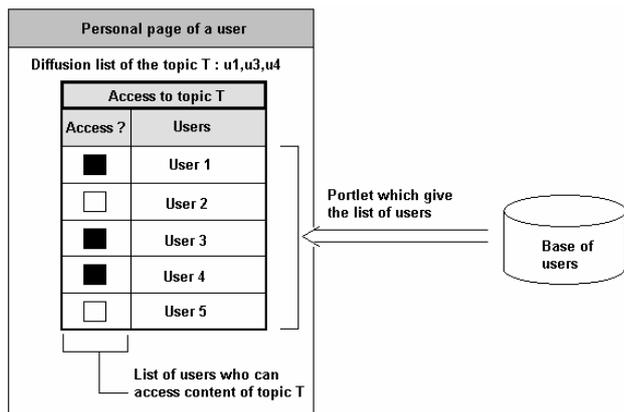

2.2.4 Communication between personal servers

Two personal servers need to communicate when a user gives an access to another user or when a user watches the topics received from other users. We deal with both these cases as follows.

2.2.4.1 A user gives an access to another user

First, the server adds the new user that receives the access to the distribution list of the topic. Secondly, a request is send to the personal server of the new user, to add the id of the topic's owner to the list of contacts.
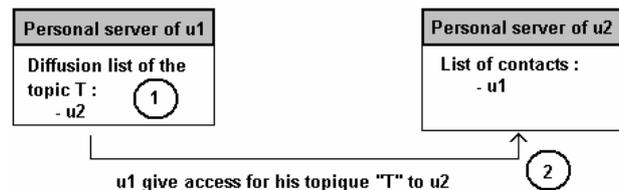
2.2.4.2 A user displays the topics received

When a user wants to see the topics he receives, his server needs to request the servers of those users that are in his list of contacts. Again we use portlets to display this result.
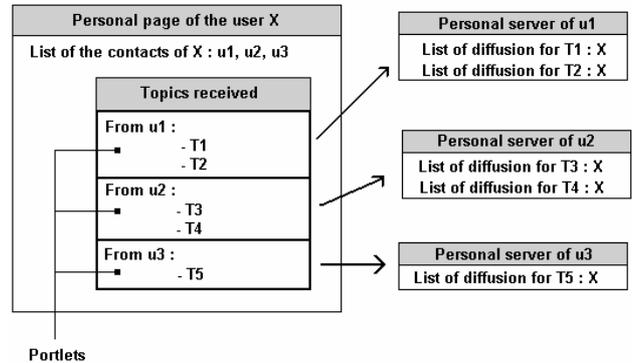
To display all the received topics, the server calls a portlet on each server of every user on the list of contacts. Each portlet returns a list of topics that the user can access.

2.2.5 Displaying received references

The page that displays received topics contains a 'Update' button. A click on this button throws a batch process that sends a request to all the personal servers of the users who send topics. This request gets all the references accessible by the user. These references are added to the XML data file. In their list of topics, they contain uniquely those topics accessible by the user. These are also added to the XML data file. This process is thrown periodically (for example once a day).

To make a difference between an owner's data and received data, the received data are added to the data file as a different type of entry (reference and otherReference, topic and otherTopic). The two types 'otherReference' and 'otherTopics' have one more attribute than reference and topic, namely "author". This attribute is the identifier of the user who sent this reference or topic.

Each entry of the XML data file has also a date attribute and an attribute specifying if this entry is related to an information creation, modification or deletion. With these attributes, it is possible to know if a received data item has been modified or deleted since the last update. Therefore, if data is already in the data file, it is not necessary to add it.

Each user manages her own topics and reviews that are stored on her personal server. Therefore no conflicts should be detected between copies of information. Anyway, if such conflict appears, solutions exist to manage them using the stored information in the XML file [5]

To display references received by a user on a particular topic, it is necessary to request the server with the identifier of the topic. To respond, the server searches all the otherReferences of the XML data file that contain this topic in its list of topics.

3. SEARCHING THE WEB OF PEOPLE

The WWW would not be very useful if search engine would not exist. It is certainly the same for the Web of People. Users could not easily break out of the network of their local neighborhood and connect to other interesting users that may provide appreciated information.

As we said previously the Web of People is the dual space of the web. Thus, a search engine of the Web of People is used to find people instead of pages. Thus, the goal of the search engine is not to propose a user to see a page but to add another user to the distribution list of one of his topics. Another use of this search engine is to help users to accept to receive information from another user.

3.1 Using the search engine

A user, let's call him u, will use such a search engine in order to find someone to send him new and interesting information about one of his topics. But in order to satisfy that query of user u, the search engine can not oblige another user, let's call him v, to add user u to one of his topic's distribution list that would really interest user u. Let us call v1 this topic that may be of interest to user v. The only action it can recommend to user u is to add user v to one of his own topics' distribution list that would interest user v. Let us call u1 this topic of user u. By doing so, user v will receive a proposition to receive information from user u about topic u1. Then user v can also use the search engine to ask him about his interest to receive this information. The search engine will then recommend user v to accept and to add user u to his topic v1 that interests user u.

One feature of this idealistic protocol is that it should stimulate cooperation and information production. To receive information from other users a user must have interesting information in order to be recommended. If he does not want to wait too long for a user to accept to send him information, he has to initiate the cooperation and needs to have appreciated information to send. Some analyses have shown that the quality of a cooperative exchange service is under threat from the number of free riders in the service [6]. Free riders are users who mainly consume information but produce very few. We think that this way of engaging cooperation should reduce free riders in the Web of People.

3.2 A search engine as a specific node

Our search engine can be seen as a specific node in the Web of People. This node has only incoming links coming from users who decide to add a search engine in one of their topic's distribution list. A user that decides to do so thus wants the search engine to find relevant contacts to receive information about these topics. But to receive information the user must also accept to send information in his topics. Then he has also to put the search engine in the distribution list of the topics from which he also agrees to send information to some users. Receiving the information being in the user topics, the search engine is going to compare the topics of multiple users. When it finds similarity between two topics belonging to two users it will recommend each of these users to add the other one to the distribution list of his topic. Search engines being nodes of the Web of People, many of them can be integrated in the network. It is really easy for a user to choose multiple search engines according to the topics he wants a search engine to find some relevant contacts for. Using multiple search engines also prevents a specific search engine to collect too much information about a specific user.

Without users' agreement the search engine does not receive any information. Because users can easily change the search engine they want to use, search engine operators have to respects their users. For example if a search engine recommends a user to receive information from contacts that turn out to be

unappreciated, for example from spammers, the user can decide to stop sending information to this search engine.

3.3 The search engine at work

Being in the distribution list of some user's topics, the search engine can access to the references included in those topics. Then the search engine is going to compute a similarity measure between topics of different users in order to propose them to send information to each other. The goal is to find a topic u1 belonging to user u similar to a topic v1 belonging to a user v in order to propose user u to add user v in the distribution list of his/her topic u1.

One classical similarity measure between topics could be based on the similarity measures of the WWW resources referenced in the topics. But those similarity measures are based on automatic content indexation and such technologies can not take into account subjective point of views. Moreover they are not always sufficiently efficient for content like images, music or videos [7]. Because people appreciate to share such contents, we think that those kinds of contents will largely navigate through the Web of People. This remarks leads us to choose another approach based on collaborative filtering techniques using only the URLs in the topics as objects to be compared in order to find similarities. But the WWW is huge and contains billions of URLs, thus the intersection between the URL sets that are referenced in topics might be more often than not empty.

To deal with this problem we first use a loose similarity measures between URLs. A very simple measure is computed as follows.:

If the URLs i and j are the same then $s_{ij}=1$
Else If the URLs point to the same repository then $s_{ij}=0.5$
Else if they have the same host then $s_{ij}=0,25$
Else 0

Of course this measure has many limitationd. Its main advantage is that it is very simple and efficient to compute.

Second, in order to find URL sets that cover a large number of URLs we use the Open Directory Project repository.

The search engine computes similarity between topics when it detects that two topics have URLs that are equal or similar to URLs classified in the same ODP category.

First the search engine computes a similarity measure between each topic and each ODP category. Like others do, this similarity measure is base on URLs co-citation analysis [8]. It is computed according to the formula:

$$S1(t,c) = \frac{\sum_{i \in t} \max_{j \in c}(s_{i,j})}{n \times m}$$

where :

t is a set of URLs in a topic
c is a set of URLs in a ODP category
$s_{i,j}$ is the similarity measures between URLs I and j
n : is the number of URLs in the C category of ODP
m: is the number of URLs in the topic T

In order to take into account the hierarchy of ODP categories and user's topics, a topic or a category includes all the URLs in all its sub topics or sub categories. We assume that the similarity measure S1 has its highest value when the category C and the topic T is not too specialized with too few URLs included in the category and in the topics or too general with too much URLs in

both that increases the scaling factor n x m and thus reduces the value of S1.

Then the similarity between two topics T1 and T2 corresponds to the result of the following formula:

$$S2(t1,t2) = \frac{\sum_{c \in ODP} S1(t1,c) \times S1(t2,c)}{\sqrt{\sum_{c \in ODP} S(t1,c)^2} \times \sqrt{\sum_{c \in ODP} S(t2,c)^2}}$$

where ODP denotes the sets of all the ODP categories.

Of course the search engine only computes the similarity between topics that do not belong to the same user.

This expensive process is made off line like the indexing process of WWW search engines. Later, the similarity measure between topics is used to satisfy user queries.

When a user u asks the search engine about new contacts about one topic u1, the search engine searches a topic v1 that belongs to a user v and that satisfies the following constraint:

S2(u1,v1) =Max S2(u1,vi) for all topics vi of user v

By doing so the search engine selects a user v that is likely to appreciate to receive information from the user u in his topics u1 and should accept to put user u in the distribution list of his topic v1.

The search engine will thus answer to user u by recommending him to add user v in the distribution list of his topic u1.

If the user u does not specify the topic u1, the search engine computes the recommendation for each topic of user u it receives. Then it ranks the recommendation according to the value of S2(u1,v1).

When a user asks the search engine if he should accept to receive information in a topic u1 from user u2, the search engine computes its answer according to the topic v1 that maximize S2(u1,vi) for all topics vi of v, and the value of S2(u1,v1).

If this value is smaller than a threshold defined by the user or by the search engine administrator then the search engine indicates it does not recommend accepting. If this value is greater than this threshold it indicates it recommends accepting.

3.4 Improvements
Many improvements can be foreseen for this basic search engine.

First if could take into account the fact that topic v1 might only contain URLs which are already in the topic u1 and thus topic v1 is not an interesting source of information. This feature would break the symmetric property of the similarity relation S2 between topics. Then if topic v1 is an interesting source of new information for topic u1, the search engine would have to find other topics u2 belonging to user u and v2 belonging to user v so that u2 is an interesting source of information for the topic v2.

Next, if for example user v is an expert in one subject he might not be interested in information provided by user u on that subject but he might accept to send him information on that subject if user u is able to send him information about another subject that he appreciates.

The search engine could also manage reputation information about topics and their users if users accept to inform the search

engine about the topics they accept and want to continue to receive.

The search engine has a partial view of the graph of the Web of People. Structural analysis on that graph could be made to rank topics and users using social network analysis techniques in the same way of the PageRank algorithm used by google [9] [10]. Such structural analysis could produce indicators of the popularity of a topic but also could detect influences of one topic on another. For example if a user u already receives information from one topic x similar to topic u1, and user v also receives information from topic x similar to his topic v1, it would not be interesting for user u to receive information from topic x, or user u should stop receiving topics v1 and start receiving information from topic x.

4. PRELIMINARY EXPERIMENTS
The first experiments with the Web of People are taking place in our France Telecom research laboratory. A pilot group of about seventy people includes engineers, researchers and PhD students. Their activities are very diverse, some working on operational services and some working on advanced research. It is clear that these users use different topic taxonomies to organize their references.

The early adopters of the Web of People were the PhD students. They use it to manage their references. They accept to share this information with researchers they are working with. These researchers appreciated to get access to selected sources of information in their research domain. Then they categorized these sources in their own taxonomy that were more similar to the operational activities' domains. This new categorization allows engineers working on operational services to discover new relevant sources of information in their own domain of activity.

Although such scenarios show the interest of users for the Web of People, this experiment also helps us to identify the source of resistance from the users to use it. The main problems were in the interface design and it's integration in the tools that people were already using everyday. For example, users asked us to integrate the Web of People access in their e-mail client. We have achieved this by sending an e-mail to a user when new reviews were associated to topics he is able to access. The e-mail contains a copy of his personal access page with active hyperlinks that lead the user to use the web browser.

In the near future we plan to conduct other experiments: One in all France Telecom R&D, and another one within the teaching community by integrating the Web of People in a web portal used by more than 400 people including college teachers, pupils, and their parents [11].

5. RELATED WORK
The Web of People proposes multiple facilities to users to manage their own personal information about resources on the WWW, to communicate with others and to exchange and share information cooperatively. Many systems have been developed to support these functionalities but few integrate them all in a single architecture.

5.1 Bookmarks management
Bookmarks managers exist since the first WWW browsers were released. The main difference with browser based bookmark managers is that with the Web of People all the information is stored and managed in an online server. There are two advantages of this network exchange solution. First, the user can access

bookmarked information from any terminal using a plain HTML browser. Increasing mobility and the availability of multiple connected mobile devices are becoming crucial. Second, this information can be accessed and shared by several users. Other solutions propose on-like bookmarks access but they are globally centralized and do not provide facilities to control the sharing of information.

The Cockpit system deals with the problem of personal bookmark sharing [12]. However, Cockpit distinguishes between community and personal vocabulary, forcing users to add more information. This system is also globally centralized.

5.2 Mail system

The most convenient media used to communicate through the information network is the mail system. Unlike general mailing systems, the Web of People is dedicated to sending information on information. We think that this specific kind of content needs a specific media to be distributed. The reasons are the following. First, several pieces of meta-information can be about the same document and specific treatments can then be applied in order to aggregate and filter them. For example when a review is seen or deleted by the user any new review about the same document will be hidden; and all reviews about the same document are displayed grouped together showing the multiple topics the user can access that have been associated to the document. Second, the documents described by that meta-information may require availability of time, bandwidth, or display size to read it. This specific media is consulted when these appropriate resources are available.

Third, exchanged multimedia documents are increasing in size while, in contrast, mobile devices are becoming smaller. By developing media that distribute documents by sending meta-information and URLs, we want to encourage users to store documents on http servers and send URLs instead of sending huge attached files by email and copying them in each recipient's mailbox. Therefore, the disk space of mailboxes could be more efficiently used and the time to download communication messages significantly reduced. Users download voluminous contents only if they need it and when they have the appropriate connection. We are aware that users have to manage and control the accessibility of online documents on the http servers to the appropriate users, but easy solutions can be developed (see for example shared document functionalities associated with group or mailing list services on main Internet portals).

Finally, meta-information is a key issue for the famous Semantic Web and needs appropriate services to stimulate their production and use [13]. We believe that the Web of People might be a medium that will produce useful information for the semantic web.

5.3 Information sharing

Although the Internet was presented as an exchange and share support and although it has allowed the development of systems such as:

- (1) newsgroups,
- (2) mailing lists,
- (3) knowledge sharing systems [14][15] or
- (4) peer-to-peer systems [16].

These collaborative systems still raise certain problems.

With newsgroups (1), the users receive a great amount of information because of the number of participants. The quantity of information could hide the interesting one. The number of participants makes it difficult to remember interesting people and to locate experts. The access to the published messages is public. As a result, the user who sends a message does not see the people who can read it. There is no control in the visibility of their publications. This is the typical source of spamming. This tendency seriously threatens the use of these community services.

With the Web of People we try to help the user to manage a few number of contacts in order to develop personal relation ships. The search engine should help the user to choose the most appropriate contacts according to his need and the availability of new users information.

Controlled sharing systems (2), like mailing lists, partially solve the problems listed above. Mailing lists allow the definition of the receiver of a sent message. However, they still contain inconveniences. For example, it is difficult for a user receiving messages to filter messages according to the emitter. He often uses the emitter's address (because he/she does not see the mailing list's name to use it as a filtering criterion) as a filtering criterion. All the messages from this user are then filtered, whatever the subject. This can be a shame. When a list is personal, only the mailing list creator can send messages to this list. When a mailing list is public, everybody can send a message to this list, but if a message interests two groups of people corresponding to two different mailing lists, if one user belongs to both, he/she will receive this message twice.

When accessing to his personal server on the Web of People the user can see only review about new resources. He can decides to not receive information about one topic from a specific user. Information being tagged and associated to uniquely identified resources (URLs) the Web of People allows many filtering possibilities.

Notice that, whatever the privacy of lists, when we manually manage distribution lists, we only insert the people we know. It is true that users can sometimes ask to be registered. Nevertheless, how could a user know about an existing distribution list? Moreover, how can he/she know if the contents could interest him/her if he/she has no access to these contents already?. The goal of the search engine of the Web of People we proposed in this paper is to solve this problem by recommending user to add new persons to their distribution lists.

Knowledge sharing systems (3) store information to be shared among users. Information is usually classified in a collective taxonomy. Users can not access to information according to their own criteria. They do not control the visibility of information. Without such control, some users are afraid and intimidated to participate. Moreover, those systems are generally centralized.

In the Web of People each user manages his own taxonomy and controls who can see his information. Moreover, the architecture we propose in this paper is entirely distributed.

Finally, new exchange systems are based on "peer to peer" (4) technologies. Their main problems are, the uncontrolled information search, the impersonality of the exchanges (although sometimes anonymity is appreciated) and, for most of them, the need for a specific client software installation. Using the Web of People requires installing a unique personal server instead of installing a client on each user terminal.

Some other systems also try to handle some the problems of

finding relevant contacts and filtering information using collaborative techniques[15]. However, this filtering process is limited by users' motivations to supply comments about the messages. Even if at the beginning the user is motivated to supply this information, the perceptible return on the quality of the obtained filtering is often long and can be disheartening. With the search engine we have presented we use the ODP repository to cover a large part of the URLs that could be reviewed with users topics and to find rapidly similarities between users profiles and to be able to recommend rapidly something to users.

We also observe that, whatever the exchange system, problems also arise from the user's behavior [6]. One of the most observed phenomena is that more users are consuming rather than producing information. Consequently the quality and robustness of the system are dependent on these producers. Moreover, these producers are powerful and can use this power in inadequate ways – like rumor distribution, advertising, etc.

The way of using the search engine we propose should stimulate a cooperative behavior of users. We are aware that more improvement needs to be done. We have proposed some ways of progress. If users allow the search engine to know who they accept to receive information from, the search engine should be able to regulate the connection between users in order to limit the power of some users.

5.4 Recommender systems

Using any kind of relevance feedback information for computing users profiles and making correlation between them have been extensively studied for building recommender systems [17]. We have already indicated how we deal with one of the main problem of these systems, which is user's motivation to provide feedback. Our search engine works as a recommender system dedicated to propose some match making between users. In that sense, this search engine can be similar to expert location systems. A detailed bibliography of such systems can be found in [18]. But our use of such "users matchmakers" is quite different. Instead of proposing to a user to receive the most interesting information from experts we propose to him to send information to any users who may find his information interesting. We suppose that this user will then receive information back from these users if they really find the received information interesting In such a way, we want to stimulate the production and diffusion of qualified information. We are aware that our system is thus more adapted for building "expert groups", each groups having a its own level of expertise, instead of promoting expert knowledge diffusion through large audience. We think this approach preserves experts from being overloaded and produces groups with more homogenous background and expertise. In order to enter to a new expert group, eventually with a "higher level of expertise", a user must have information which one member of this group might found interesting. The search engine will then propose him to send this information to this member. Then, this member might introduce the user to the expert group by sending back to him information from this group. All these assumptions will be checked during our experimentations.

5.5 Social networking

Our work is close to works presented by Prabakar Raghavan about social networking [19]. The web of people is a specific example of using social networks in order to find information adapted to personal needs. We created a specific environment in order to stimulate the production of such a network and the exchange of information through it. The idea of using people as network nodes can also be found in the Pollen project conducted at Xerox XRCE research center [20]. In their project, information is directly exchanged through user's devices whereas we use the WWW for exchanging information.

6. CONCLUSIONS

The Web of People, like the Semantic Web, should be seen as an augmentation of the current WWW, a kind of meta-Web. In our analysis of Web paradigms we are positioning it either as a complementary paradigm to the Semantic Web, or as one of several fallback scenarios should the Semantic Web, for one reason or another, fail to materialize or to impose itself. Even if each of those paradigms can survive independently our goal is nevertheless to develop positive interactions between both of them. For example, objective and formal meta-data managed by the semantic web, and more subjective, personal and loosely constrained metadata managed by the Web of People should be easily combined to offer more personalized filtering and searching functionalities. One of the most attractive features of the Web of People, we think, is that it provides for a way to deal with multiple and personal ontologies, thus avoiding the need for users to use standardized ontologies (such standardized ontologies are of course useful for the formal and machine processable information in a Semantic Web). Moreover, it deals with the problem of mapping ontologies, not by automating this in any sophisticated way. Instead it provides a relatively low-tech solution by tapping into the intelligence and interpretation capacity of its users. One direction to explore is whether this can be used as a way to generate common ontologies in a bottom-up way. The Web of People can therefore also be seen as a first step towards a self-organizing Web, driven by its use.

We have described a architecture of the Web of People, and its current realization in SoMeOne. Further work on this architecture and on the search engine for the Web of People has been indicated in Section 3. In particular we need to establish the scalability of the system, both at the technical level and at the cognitive level for its users.

In the near future we plan to conduct such larger experiments: One in all of France Telecom's R&D, and another one within the teaching community by integrating the Web of People in a web portal used by more than 400 people including college teachers, pupils, and their parents [11].

7. ACKNOWLEDGMENTS

8. REFERENCES
[1] Stefik, MJ.Internet Dreams: Archetypes, Myths, and Metaphors. MIT Press, Cambridge, MA, 1996

[2] Berners-Lee, T., Fielding, R., and Frystyk H., 1996. Hypertext Transfer Protocol HTTP/1.0., RFC 1945 MIT/LCS, UC Irvine, May.

[3] Jalios : http://www.jalios.com

[4] Network Time Protocol http://www.ntp.org

[5] Dedieu, O. Réplication optimiste pour les applications collaboratives asynchrones. *Thèse de Doctorat de l'Université de Marne-la-Vallée. 14 septembre 2000.*

[6] Huberman B..The laws of the web, patterns in the ecology of information (2001) MIT Press

[7] Salton, G. and McGill, M.J. Introduction to modern information retrieval Mc Graw-Hill 1988

[8] Kessler M.. Bibliographic coupling between scientific papers. American documentation, 14:10--25, 1963.

[9] Brin, S. and Page, L.. The anatomy of a large-scale hypertextual web search engine. In Proceedings of the WWW7 Conference, 1998

[10] Wasserman, S. Faust, K. Social Network Analysis (Cambridge 1994): Cambridge University Press.

[11] Martel C, Vignollet L,. Educational Web Environment based on the metaphor of electronic schoolbag, ARIADNE 2002, Lyon, France.

[12] Gräther, W., Prinz, W.,.The social web cockpit: support for virtual communities. In ACM SIGGROUP Conference on Supporting Group Work, ACM Press, Boulder, Colorado, USA, 2001, pp. 252 - 259.

[13] Berners-Lee , Hendler J and Lassila O. The Semantic Web. Scientific American, May 2001

[14] Bouthors V, Dedieu O, 1999. Pharos, a collaborative infrastructure for web knowledge. In proceedings of ECDL'99. Lecture Notes in Computer Science, 1696, pp. 215-233, Springer-Verlag Inc., (September).

[15] Glance N, Arregui D,. Making Recommender Systems Work for Organizations Proceedings of PAAM'99, London, UK, (April 19-21).

[16] Peer to peer working group http://www.p2pwg.org

[17] *Recommender Systems.* Special section in Communications of the ACM, Vol. 40, No. 3; March 1997

[18] D. Yimam, "Expert finding systems for organizations: Domain analysis and the demoir approach," in Beyond Knowledge Management: Sharing Expertise (M. Ackerman, A. Cohen, V. Pipek, and V.Wulf, eds.), (Boston,MA), MIT Press, 2000. (http://citeseer.nj.nec.com/490565.html)

[19] Raghavan, P. Social Networks: from the Web to the Enterprise. IEEE Internet Computing, pp. 91-94, Jan/Feb 2002

[20] Natalie S. Glance, Dave Snowdon, Jean-Luc Meunier: Pollen: using people as a communication medium. Computer Networks 35(4): 429-442 (2001)