# A multimodal browser to navigate and search information on the Web

José Rouillard, Jean Caelen

# A multimodal browser to navigate and search information on the Web

**José Rouillard** and **Jean Caelen**

Laboratoire CLIPS-IMAG, Groupe GEOD
Université Joseph Fourier
Campus Scientifique, BP 53, 38041 Grenoble Cedex 9 - France

*E-mail : Jose.Rouillard@imag.fr*
*E-mail : Jean.Caelen@imag.fr*

**Abstract**
This paper describes a prototype of browser which uses speech recognition to navigate and retrieve information through the World Wide Web. Our multimodal browser shows how it is possible to achieve, with speech, some tasks which are classically executed by the means of the keyboard and/or the mouse. We describe the way we use natural language to help users, novices or experts: with dialogue and sub-dialogue procedures included in an interaction system, we guide the user to a man/machine conversation where he/she can tell his/her needs and goals.

## 1. Introduction

Internet and more generally the World Wide Web is growing very quickly and the number of people surfing the Web increase every day. But today we have to face a new kind of matter: classical browsers were originally made for scientific, military and professional users. Nowadays, more and more people are connected on the net, then it is interesting to see how multimodal interfaces could help them to navigate and retrieve information through the Web [BERK 91]. According to Conklin [CONKLIN 87] the two most important problems related to information access through hypermedia interfaces are disorientation and cognitive overload.

Some researchers are currently working on «How to personalize the Web» because they noticed in a recent survey that "the main problems people report when using the web are : (a) slow network or connection speed (b) not being able to find particular pages, even after they have been found before (c) not being able to manage or organize retrieved information and (d) not being able to visualize where they have been."[PITKOW 96].

Since Salton and McGill, [SALTON 83] we know that in an information retrieval system, we have to present to the user all and only the relevant information. But how can we help them when we know that «some users wanted the ocean, while others were looking for a grain of sand» ? [HARDIE 96]. We are convinced that a user/computer oral dialogue will be able to lead people to a more relevant interaction with their machine, particularly for the inexperienced users.

Some French [KABRE 95b] and American [HOUSE 95] works showed that it was possible to navigate the World Wide Web by means of the voice. This paper presents a prototype of oral browser which is not only able to control the basic commands used on the Web (back, forward, home, etc.) but can also help the users in different ways, according to the meaning of the words they use in their request [FLANK 95].

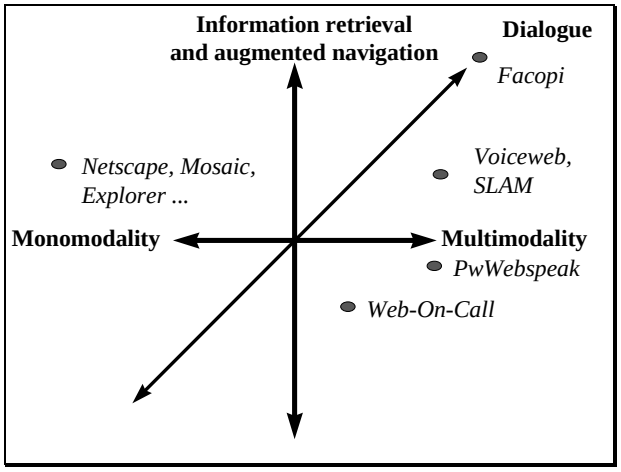


**Figure 1: Browsers position map**

## 2. Methodology

Our browser, named FACOPI[1], is both multimodal and dialogue-oriented (see Figure 1, [RHIE 96], [HOUSE 95], [HAKKHINEN 96], [KABRE 95b]). It runs on Windows 95 PC computers and goes through seven different parts (see Figure 2) which are the following:

- 1) *speech recognition*: uses the «Echotalk» word-spotting module based on a Hidden Markov Model (HMM) (developed in our laboratory [CAELEN 96], [KABRE 95a]). Echotalk runs in background and associates a string of characters (sent to the browser's window) to a keyword of its lexicon.
- 2) *semantic concept analysis*: this module associates the recognized keywords to a more general concept. The most unlikely combinations are forgotten while some others are added if they seem relevant and correlated to the initial oral request.

[1] FACOPI : Feuilleteur A Commandes Orales Pour Internet (English translation : Internet Multimodal Command Browser).

- 3) *query expressing:* prepares a provisional query according to the chosen semantic field, denoted by the concept.
- 4) *multimodal dialogue procedure:* offers a temporary query to the user and dialogues with him/her to confirm, modify or cancel it.
- *5) submission to the engine:* only by sending the query to a powerful search engine.
- 6) *multimodal answer presentation:* some information is presented on the screen (text, pictures) while some others are synthesized on the audio channel (for instance: general information on the results, titles, short abstracts, etc.).
- *7) loop and possible relevance feedback:* as we move forward or backward to take a relevant picture, here it will be possible to «zoom in» (if we want more details) or «zoom out» (if we require to include more components from the related domain). In this part, it is important to record the user's satisfaction: how and in which way the machine helps him/her. Later, it would be easier to guide other users in the same way.

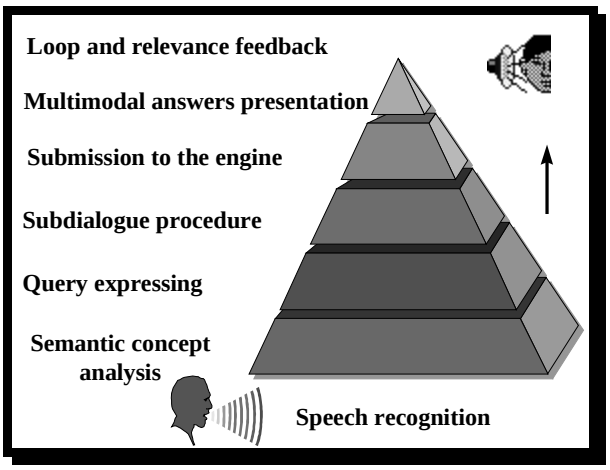


**Figure 2: The FACOPI interaction system**

**3. Dialogue with the user**

Before sending a query to the browser engine, a phase of dialogue with the user allows to refine and confirm the request: the system analyses the pronounced word and if it matches an item of the database, the interface asks for the field to use: one or several themes are attached to each word, and each theme is connected to one or several URLs. For example, if the user pronounces the word «bear» (see Figure 3), does he/she want some information concerning a star (the Great Bear, the Little Bear) or rather about an animal (white, brown, polar bear) or even something else ? In this situation where a keyword is correctly recognised but the context does not give enough data, a sub-dialogue procedure asks the user to choose among several semantic fields, according to its extensible database. In our example, the system proposes a choice between the following research domains: Astronomy, Animals. So, whatever the user's choice, the browser shows appropriate URLs. And if no choice is selected among the different themes, then the browser can submit the query to a powerful spider Web service. With a system able to identify many words in the same sentence, the ambiguity will disappear with the connection of keywords from a common semantic field. If the keywords «great» and «bear» are detected in the same speech, the system will surely understand that the relevant semantic field is «astronomy», and then, it will present a proper address on the web, and ask the user to confirm or deny this proposition.

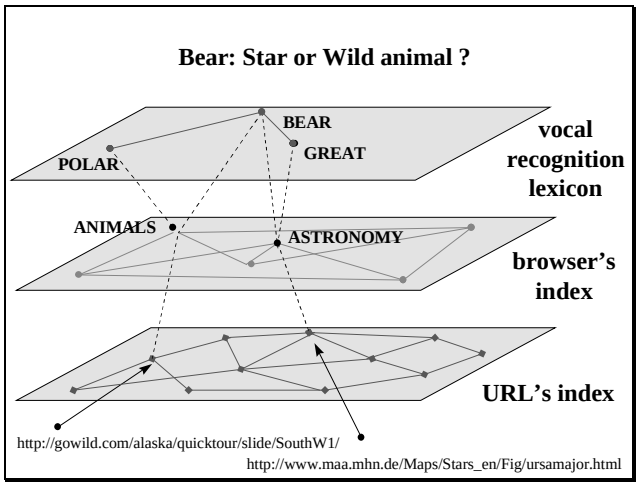


**Figure 3: Interconnected semantic networks**

Some analyses on the nature of the dialogue between humans showed that, according to the task to achieve, the kind of dialogue can be significantly different [CAELEN 97]. The strategies observed in those tasks are often among the following ones:

(a) reactive strategy: an action brings a reaction, no more, no less.
(b) cooperative strategy: there is more information exchange between the speakers, and more speech turns to be sure that the needs and possibilities of each person are understood by the other.
(c) negotiated strategy: when the goals of the speakers are not in the same direction, they have to talk together, to negotiate, and to found a solution satisfying both of them.
(d) constructive strategy: it mainly occurs when the discussion diverges and deviates from the initial main subject.

Our statistics on those kinds of strategies in some human-human dialogues indicate that 80 percent of the transactions are cooperative (see Figure 4, [CAELEN 97]). Of course, in our Human-Computer Interface, we probably have to emulate this strategy, if we want to give the user the most relevant information.

| | |
|---|---|
| Reactive strategy | 18,33% |
| Cooperative strategy | 80,61% |
| Negotiated strategy | 0,76% |
| Constructive strategy | 0,30% |

**Figure 4: Strategies observed in a human-human dialogue corpus**

| French Index | Ambiguity | Theme | Relevance | HTTP Title | URL |
|---|---|---|---|---|---|
| cours | 0.7 | finance | 23 | Paris stock exchange | http://www.fastnet.ch/NETFUND/HTML/rm.html |
| | | | 17 | New York stock exchange | http://www.cnnfn.com/resources/links/corp.html |
| | | | 6 | Indicative price (livestock) | http://www.agri.ch:80/amarkt/schlacht/schlachf.htm |
| | 0.3 | teaching | 3 | computer course | http://fillmore.univ-mlv.fr/~dr/Cours.html |
| télé | 1 | television | 15 | programs grid | http://www.oceanis.net/ak24/TV/current.html |
| | | | 8 | TF1 (channel 1) | http://www.tf1.fr/prog/cinema/d12.htm |
| | | | 6 | France2 (channel 2) | http://www.france2.fr/vos-soirees/prog-jeu.htm |
| | | | 5 | France3 (channel 3) | http://www.sv.vtcom.fr/ftv/fr3/prog/jour.html |
| ... | ... | ... | ... | ... | ... |

**Figure 5: Semantic network index of the system**

An extensible index database is used and each word of a scope is balanced with a weight according to the relevance, estimated by the user's satisfaction. The aim of this oral browser is also to help the user in a cooperative dialogue which leads to a non-ambiguous query. For each keyword of the index, there is a recorded vocal pattern to compare with the words extracted from the speech of the user. One item can be attached to one or several themes. An ambiguity rate is associated to each theme. In our example (see Figure 5) the keyword «cours» is linked to a financial or a teaching domain. The financial domain will be presented first to the user because its ambiguity rate showed that (at that very moment) 0.7 of the past query which used this keyword were in relation to a financial semantic field, and only 0.3 for the teaching one. This rate is dynamic and changes every session with the user's satisfaction. It is written for n keywords:

$$\sum_{i=1}^{n} Ai = 1$$

When the user chooses a theme, there is a sub-dialogue and the system offers one of the multiple URL recorded for a particular domain. The relevance of an URL according to the index theme measures how many times a user jumps to this page. We see in our example that «télé» was always used to talk about «television» and it was used 15 times to watch the television grids on the World Wide Web.

Concerning the meaning and understanding of the words, one of the solutions brought by some researchers, is to use multi-layered neural networks for the semantic settings and data micro-semantic analysis [ANTOINE 94]. The Digital Equipment Corporation uses the *LiveTopics* technology for its Altavista search engine on the Web: it is a visual interface which offers a thematic map as an answer to a specific request [BOURDONCLE 97].

With a concept analysis, it is possible to extend the semantic field by connecting pieces of information which have close meanings. A semantic concept analysis is used in this oral browser. We noticed that, for specific fields, the number of non-relevant links was reduced by 95% for the best scores. This improvement was due to the use of semantic fields instead of words or combination of words.

We made an experiment which showed that a French user who wants to know, via the Web, the value of a share from the Eurotunnel company expressed his query in this natural way «Je veux connaître le cours de l'action Eurotunnel»[2]. The most relevant words to compose a query, according to this question are «cours», «action» and «Eurotunnel». But in French, those words have different assertions depending on the context and the historical background.

[2] « I want to know the Eurotunnel's share value »

| Terms of the query: | Rank of the relevant link | semantic dominance |
|---|---|---|
| cours | 123 | Lesson, course of ... |
| action | 0 | To do something |
| Eurotunnel | 0 | The shuttle |
| bourse | 85 | Finance, Exchange, Thesis. |
| Paris | 19 | France capital |
| | | |
| cours+action | 0 | Lesson |
| cours+Eurotunnel | 143 | Lesson ; course of ... |
| cours+bourse | 8 | Financial market |
| cours+Paris | 3 | City of Paris |
| | | |
| action+Eurotunnel | 0 | The shuttle |
| action+bourse | 91 | Financial market |
| action+Paris | 0 | To seer / to do ... in Paris |
| | | |
| Eurotunnel+bourse | 95 | Finance |
| Eurotunnel+Paris | 0 | ? |
| | | |
| bourse+Paris | 2 | Paris Stock Exchange - CAC 40 |
| | | |
| cours+action+Eurotunnel | 0 | Lesson, course of ... |
| cours+action+bourse | 12 | Financial market |
| cours+action+Paris | 3 | Financial market |
| | | |
| action+Eurotunnel+bourse | 103 | Exchange, Financial market |
| action+Eurotunnel+Paris | 0 | ? |
| | | |
| Eurotunnel+bourse+Paris | 8 | Financial market |
| | | |
| cours+action+Eurotunnel+bourse | 9 | Financial market |
| cours+action+Eurotunnel+Paris | 54 | Exchange, Financial market |
| | | |
| cours+action+Eurotunnel+bourse+Paris | 4 | Financial market |

**Figure 6: Statistics about the rank of relevant terms**

We studied the results of the combinations of those keywords (1 by 1, 2 by 2, etc.) which were sent to the engine Altavista (see Figure 6). Alone, only one of the first three terms leads to a positive result (the 123rd link-answer of the query «cours»). The words «cours», «action», «Eurotunnel», «bourse» and «Paris» have been respectively retrieved 186234, 1670983, 944, 13437 and 409645 times by the search engine. But in fact, the semantic dominance changes with the combination of words in the query. And it is not better with queries made of two or three words together.

So, the results of this experience shows that the best way to find the relevant information according to the user goal, is to ask the engine for «bourse+Paris»[3]. In this example, with a non ambiguous query, the 2nd link instead of the 123rd offers the appropriate data (it is 98.37% better). The user is not necessarily an expert of the domain. He/she just wants some data, but doesn't know where to find it. In such a context, the system is able, via a cooperative dialogue, to asks the user on his/her intentions and offers some possible ways to go on with the navigation.

Figure 7 shows the interface of our FACOPI browser. In the speaker's speech window, the system has recognised the keyword «cours», and proposes in a multimodal way the different possible choices. The financial semantic field is presented first, because its ambiguity rate (0.70) is greater than the teaching one (0.30). Each items of this domain is displayed on the screen according to its relevant rank (in our example: Paris stock exchange first, New-York stock exchange, etc.). If the user accepts a proposed URL, the browser loads the relative page and displays it on the screen.

[3] « Stock exchange+Paris »

If the user's aim is satisfied, he/she can active a process: it will updates all the weights in the database, and increase the ambiguity rate observed in this session. Thus, next times this user (or another) employs our multimodal browser, the appropriate information will be available immediately.

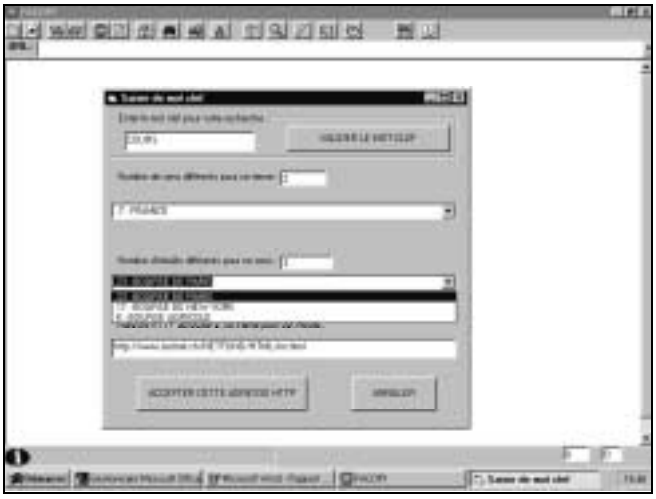

**Figure 7: After the recognition keyword, the system proposes different domains (semantic fields)**

**4. Conclusion**

In this paper we explained how it is possible to navigate on the World Wide Web by means of the voice. We presented our «Internet Multimodal Command Browser » interaction system, and showed its capabilities: not only navigate, but also search information via a cooperative dialogue between the speaker and the machine.

A kind of tool like this oral browser could be interesting in many situations: for impaired persons, for example, other means of self-orientation, navigation and information access are needed. For people who are not used to computers, it will be more intuitive to activate a browser by means of natural language. For trained users, it will be a more pleasant way to navigate. Moreover, the recognition of spoken commands will meet the need for a more user-friendly way of interacting with the system. Instead of having limited and «artificial language» commands, a more natural way of interacting will be possible for everyone with such tools.

Now, we are preparing an oral browser based on a word-spotting speech recognition system. This will allow us to catch not only one word of the sentence pronounced by the user but several ones. And then, using a large semantic concept analysis, it will be easier to understand what the user wants to do and how it is possible to help him/her, according to the previous stored results. It will lead us to a new kind of vocal recognition: a concept spotting.

**References**


[ANTOINE 94] ANTOINE, J-Y., *Coopération synthaxe-sémantique pour la compréhension de la parole spontanée.* Thesis: INPG Grenoble, December 1994.

[BERK 91] BERK, E. & DEVLIN, J., *Hypertext / Hypermedia handbook*, McGraw Hill, 1991.

[BOURDONCLE 97] BOURDONCLE, F., *LiveTopics: recherche visuelle d'information sur l'Internet*, Dossiers de l'Audiovisuel, La Documentation Française (to appear).

[CAELEN 96] CAELEN, J., KABRE, H., DELEMAR, O. & PIARD, J., *Reconnaissance robuste de la parole: vers l'utilisabilité*, Jep'96, pp. 325-329, Avignon, 1996.

[CAELEN 97] CAELEN, J., ZEILIGER, J., BESSAC, M., SIROUX, J., PERENNOU, G., *Les corpus pour l'évaluation du dialogue homme-machine*. Premières Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue de l'AUPELF-UREF, Avignon, Avril 1997.

[CONKLIN 87] CONKLIN, J., *Hypertext : an introduction and survey*, IEEE Computer, pp. 17-41, September 1987.

[FLANK 95] FLANK, S., *The role of natural-language processing in multimedia*. ACM Computing surveys, volume 27, N°4, December 1995.

[HAKKHINEN 96] HAKKHINEN, M., & INGRAM, R., *pwWebspeak*
http://www.prodworks.com/pwwebspk.htm

[HARDIE 96] HARDIE, E., *A grain of sand or the ocean ; User aims in search engine interactions.* Fifth International WWW Conference - Poster Proceedings, INRIA/CNIT, Paris La Défense, Mai 1996.

[HOUSE 95] HOUSE, D., *Spoken-Language Access to Multimedia (SLAM): A Multimodal Interface to the World-Wide Web*, Masters Thesis, Oregon Graduate Institute, Department of Computer Science & Engineering, April 1995.
(http://www.cse.ogi.edu/CSLU/publications/papers.htm)

[KABRE 95a] KABRE, H., *Echo User Guide*, CLIPS report no 1, 1995.

[KABRE 95b] KABRE, H. & CAELEN, J. *VoiceWeb : une interface vocale pour l'accès au réseau Internet*, IHM'95, Grenoble, 1995.

[PITKOW 96] PITKOW, J.E., and KEHOE, C.M. *Emerging trends in the WWW user population*. Communications of the ACM 39,6, pp. 106-108,1996.

[RHIE 96] RHIE, K., *Netphonic's Web-On-Call Voice Browser*. Fifht International WWW Conference (Workshop Web Accessibility for People with Disabilities ). INRIA / CNIT, Paris La Défense, Mai 1996.

[SALTON 83] SALTON, G, McGILL, M.J., *Introduction to modern Information Retrieval*, McGraw-Hill, 1983.