



HAL
open science

HALPIN: A Natural Language Information Retrieval System For a Digital Library on the World Wide Web

José Rouillard

► **To cite this version:**

José Rouillard. HALPIN: A Natural Language Information Retrieval System For a Digital Library on the World Wide Web. Webnet'99, 1999, Honolulu, United States. hal-02441227

HAL Id: hal-02441227

<https://hal.science/hal-02441227>

Submitted on 15 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HALPIN : A Natural Language Information Retrieval System For a Digital Library on the World Wide Web

José Rouillard

Laboratoire CLIPS-IMAG, Groupe GEOD, Université Joseph Fourier –
Campus Scientifique, BP 53, 38041 Grenoble Cedex 9 – France
Tel : (33) 4 76 63 56 51, Fax : (33) 4 76 63 55 52, E-mail : jose.rouillard@imag.fr

Abstract: Designing computers with the ability to talk and understand a natural language (NL) conversation is a major field of research. We have developed the HALPIN¹ system to implement our multimodal conversational model for information retrieval in a digital library via the World Wide Web. This dialogue-oriented interface allows access to INRIA's² database, on the Internet, in a natural language mode. The inputs from the user, could be either oral sentences or mouse/keyboard actions. The results of the first experiments show that the HALPIN system provides interesting dialogues, in particular with beginners. It gives oral responses via standard browsers, for a more natural human-machine interaction according to the user's goals and skills. This leads to successful information retrieval, while searches with the original user interface (traditional web form) failed.

1. Introduction

Retrieving relevant information on the World Wide Web is not an easy task. As [HARDIE 96] observed, while seeking a document, "*some are looking for the ocean and some others for a grain of sand*". Pitkow and Kehoe remarked that these difficulties have four possible origins: "*The main problem 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]. We believe that with the exception of the connection speed, the other problems are mainly related to human-machine interaction. In addition, according to Conklin [CONKLIN 87], the two most important problems related to information access through hypermedia interfaces are disorientation and cognitive overload. To attenuate these problems, our approach is to integrate "human skills" to the actual systems. Indeed, natural language interfaces excel in describing entities that are not currently displayed on the monitor. These strengths are exactly the weaknesses of direct manipulation interfaces, and conversely, the weaknesses of natural language interfaces (ambiguity, conceptual coverage, etc.) can be overcome by the strengths of direct manipulation [COHEN 92]. "*Direct manipulation and natural language seem to be very complementary modalities. It is therefore not surprising that a number of multimodal systems combine the two*" [CHEYER & JULIA 95]. With the growing number of Internet users, efforts to make computer interfaces that are more simple and natural become increasingly important. We think that multimodal and natural language solutions for the World Wide Web should help people to accomplish their tasks more effectively than with traditional tools. The lack of real human-machine dialog corpus, especially in French, leads us to propose a usable incremental system for the Web rather than "Wizard of Oz" experiments. In the following, we describe the modelling and implementation of our system, designed for an information retrieval task in a large digital library of the Web.

2. HALPIN system description

Our work is based on the results of the ORION project³ which concerns new multimodal technologies for Web navigation and information research [ROUILLARD 97a], [ROUILLARD 97b]. The HALPIN system is an incremental one : after each connection period, the dialog files are analysed ; the missing or

[1] Hyperdialogue avec un Agent en Langage Proche de l'Interaction Naturelle

[2] Institut National de Recherche en Informatique et Automatique (83297 documents available)

[3] <http://www.gate.cnrs.fr/~zeiliger/Orion99.doc>

misunderstood vocabulary is added or modified. This way, the system can be improved with each new version. It uses Xerox's morphological tools [GAUSSIER & al. 97] to convert the user sentences to a canonical form which may be analysed more easily by the HALPIN concept detection module. Figure 1 shows the interface of the HALPIN system for an information retrieval task on the World Wide Web.

2.1. Dialog manager cooperative model

There has been a great deal of research on information retrieval, but very little in which the NL plays an important role on the Web. Most of the classical user interfaces and search engines try to improve the efficiency of the search task with better indexing and retrieval methods. In such models, the human-machine interaction is limited to exchanges of the type: query => database access => reply. As information seeking and retrieval are interactive processes, we believe (as do [BATEMAN, HAGEN & STEIN 95]) that providing a flexible and cooperative human-machine dialogue is a complementary means of improving information retrieval systems. Inspired by the works of Brun [BRUN 98], our dialog manager uses a two-step algorithm of concept recognition which leads to an understanding of the user's queries.

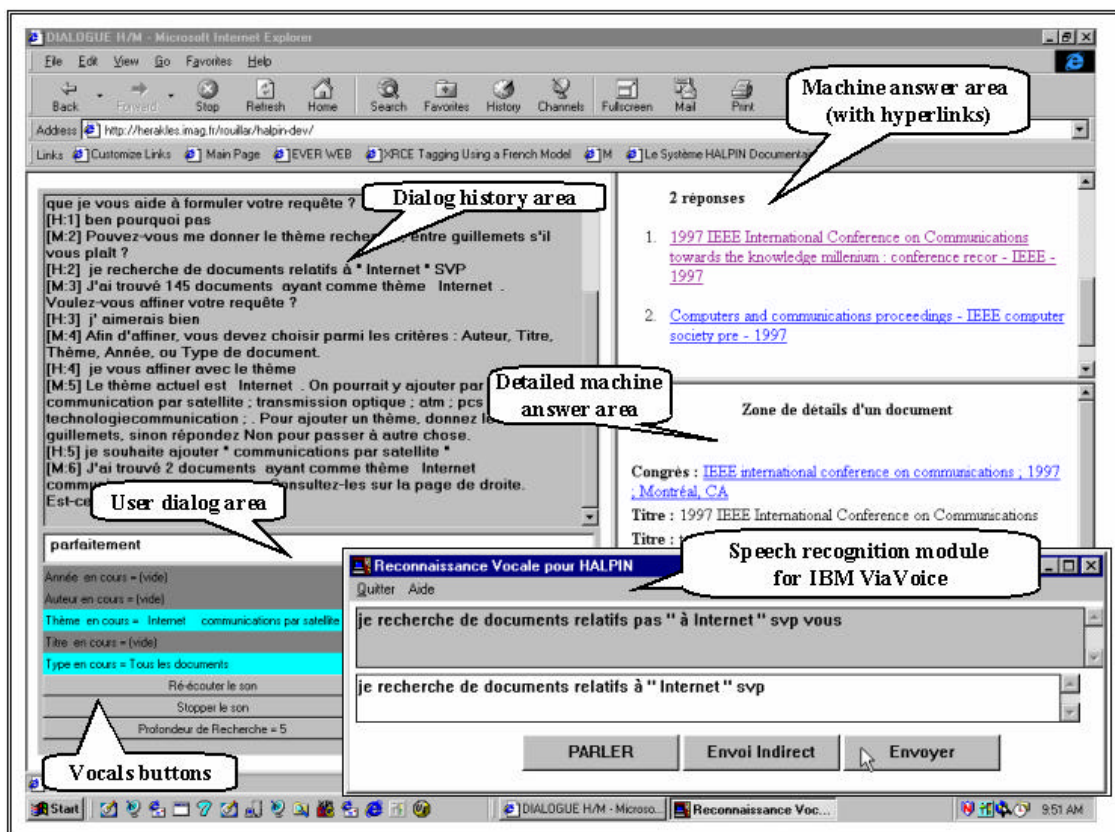


Figure 1 : The HALPIN system interface in a World Wide Web browser

An intelligent conversational system must be capable of adapting itself to the user's goals and capabilities, interpreting speech acts within the context and negotiating ambiguous information using a natural language interface [STEIN & al. 97]. In our previous papers, we have also shown that it is possible to gather interesting human-machine dialogues on the Web without using the Wizard of Oz strategy [ROUILLARD & CAELEN 98], [ROUILLARD 98]. Having in mind these important observations in order to show that NL dialogue systems may improve interaction quality, we proceeded to create an interactive search and navigation environment (HALPIN) to incorporate adaptability and conversational capabilities to an existing digital library information retrieval system. Our goal was to propose a system which not only responded to the user sentences, but also proposed related information (similar authors or keywords) depending on user needs. This is why at the beginning of the *hyperdialog* [ROUILLARD 99], with this system, we have to determine the user profile (novice or expert) and her aim. The COR (Conversational Roles) model of

[STEIN & MAIER 95] proposed typical *Ideal* and *Alternative* dialogue sequences (cycles). For example, a dialogue between the information seeker A and the information provider B, can be formalised as :

Dialogue (A,B) → request (A,B) + promise (B, A) + inform (B,A) + be-contented(A,B)
Dialogue (A,B) → offer (B,A) + accept (A, B) + inform (B,A) + be-contented(A,B)

In the same way, our model is a kind of conversational roles and tactics (COR) model augmented with the knowledge about the user and her aims, so the model can react according to the user profile and the task in progress. We propose, for a finalized and cooperative defined task, to follow the rule:

[Profile]. [Goal]. [Speech Act]. [Concepts]. [Task] → [Reply]. [Justification]. [Suggestion].

A user profile (for instance novice or expert) is determined at the beginning of the session according to whether the user wants (or does not want) some help from the machine. This profile can change if the relevant elements are detected by the machine during a session. The goal of the user, which can be “finding an already known paper”, “searching an unknown set of books”, “discovering the site”, etc., is also an important element used by the machine for the orientation of the dialogue. Speech acts and concepts are determined for each sentence given by the user. It is not only a keyword detection. For example, the concept recognition module is able to understand that “the man who wrote this book” has to be considered as the concept “the author”, or that “why not” means “yes”, which is not possible with a simple keyword matching. This method adds a robustness to the system in the understanding of some “imperfect” sentences given by the speech recognition module. For example, in Figure 1, we can see that the speech recognition module gave the sentence “H4 : Je vous affiner avec le thème” (I you refine with the theme) instead of “Je veux affiner avec le thème” (I want to refine with the theme). This is not really a problem for the dialog manager which understand the concept *Refinement (theme)* and which did not detect any negation in the sentence : the result of this analysis is a request to the digital library with the new criteria, that leads to the cooperative answer of the machine that proposed to the user some themes close to the first one (see M5 on Figure 1).

This way, the machine prepares a reply, a justification of this reply, and a cooperative suggestion if possible. The concepts database is divided into different files, according to the type of concepts. Indeed, certain concepts are common to all possible tasks (acceptance, refusal ...) and others are specific to a particular task. If a sentence is ambiguous, even when the goal of the user is known, the system proposes alternatives choices. For example, this French sentence : “je veux un livre d’algèbre de Boole” can be interpreted in two different ways : either the user wants a book talking about Boolean algebra or the user wants a book written by Boole. In the digital library that we used on the World Wide Web (INRIA), the first interpretation (theme=Boolean algebra ; author=?) gives 100 responses, while the second (theme= algebra ; author=Boole) gives only 1 response. So, we think that, rather than searching the database with an uncertain query, it’s better to resolve the ambiguity in a cooperative way. The dialogue manager allows not only entries related to the current task, but also about the interface (screen, sound, speech synthesis) and system responses (called meta-information). The system tries to understand, according to the context and found concepts, if the user is speaking about the task (ex “The author is Turing”), about the interface (ex “I can’t see anything on this screen”) or about the meta-information (ex “Why do you ask me that ?”).

2.2. Speech synthesis and recognition on the World Wide Web

The HALPIN system was developed using C, Java, Html and Perl languages. The users can hear the system responses thanks to software (Elan Informatique speech engine⁴) installed on our Web server which synthesises textual responses to a specific audio file⁵. This file is sent to the browser, and then played by the Java applet. We believe that for speech recognition on the World Wide Web, two different solutions are possible : a remote or a local solution. The first one uses a speech recognition server, so the user does not have to have its personal speech recognition software. An application must be installed on the client machine to catch the user’s voice. It can be done in a free hand way, by using the vocal energy level to determine the beginning and the end of the sentence. The sound is then sent to the server and an ASCII string is received as the answer by the client. The second possibility is that every user installs speech recognition software on her computer, connected to the HALPIN system browser window. Each method has advantages and

[4] <http://www.elan.fr>

[5] Sun Java/Web format : 8000 Hz, 8 bits for our needs

drawbacks : for the user, the first solution is the cheapest, because no software is needed but it is also the slowest, because the voice have to travel on the Web to be recognised and interpreted. So the gain brought by the “natural interface” could be lost if the speech file takes too long to travel on the network. For the moment, we have implemented software that uses the IBM ViaVoice speech recognition. It’s a push-to-talk solution that will shortly be replaced by a free hand tool.

3. First Results

We have tested our system with different types of users (novice, expert, with particular needs or not, etc.). Compared to the INRIA’s traditional Web form, HALPIN gives some relevant information according to the needs and skills of the users. The following dialogues illustrate how the machine interacts with the user to determine her needs, asks her whether these have been met satisfactory and proposes alternatives when the results are not satisfying.

129.88.41.73 on Mozilla/4.0 (compatible; MSIE 4.01; Windows NT) the 18:1:99
15:20:44:M=What do you wish to do ? 1: Find a known document, 2: Seek a set of documents, or 3: Discover this site?
15:20:44:H=1
15:21:05:M= Find a known document. OK. Do you want me to help you to formulate your request ?
15:22:38:H=yes
15:22:42:M=Do you know the name of the author of the document you are looking for ?
15:24:42:H=I believe that it is Krakoviak
15:24:56:M=I did not find a document with Krakoviak as the author. Do you want to modify your request ?
15:25:06:H=okay
15:25:13:M=Make a choice among the criteria: Author, Title, Topic, Year, or Type of document.
15:25:41:H=I would like to modify the name of the author please
15:25:55:M=Give me his name with a capital letter. You can also erase this author. Or do you me to look for authors close to Krakoviak ?
15:26:38:H=hum, yes
15:27:56:M=The authors close to Krakoviak are krakowiak, sacha; krakauer, Lawrence ; Give a new name or say No to move on.
15:28:46:H=Well, I will test with Krakowiak this time
15:29:04:M=I've found 15 documents with Krakowiak as author. Do you want to refine your request ?
15:29:24:H=Of course
15:29:38:M=To refine, you must choose among the criteria: Author, Title, Topic, Year, or Type of document (...)

Capture 1 : An example of cooperative and relevant human-machine dialogue with HALPIN

In Capture 1, the user says that she is looking for a document she already knows ; and she accepts a help given by the machine. She believes the name of the author is Krakoviak, but this name gives no answer. The machine asks for a modification of the query, and proposes to choose among : author, title, theme, year or type of document. The user chooses to work on the author’s name, and the machine proposes some names close to Krakoviak. Finally, using the relevant name of the author (Krakowiak with a W), she finds 15 documents, and the dialogue continues to refine those results.

From a critical point of view, this dialog is not optimum because the machine should have offered the near matches in 15:24.56 already. The principle of cooperation according to the maxims of Grice [GRICE 75] states contributions to the conversation should correspond to the waiting time of the other interlocutors and according to the status report and goal of the conversation. For that, four maxims are retained: 1) Maxim of method: to be clear, avoid obscure expressions, to avoid ambiguity, to be concise and to be ordered. 2) Maxim of relevance: to be relevant. 3) Maxim of quality: not to say what you believe is false, not to say something for which you have insufficient evidence. 4) Maxim of quantity: to make the contribution informative, but not more than is necessary. As we will see later, access to the digital library is an important

part of the time needed in the interaction. Searching for the authors close to the given one took, in our example, 1 minute and 18 seconds. If the user knows what the machine is doing during this time (as in our example), this latency is relatively acceptable. On the other hand, if the machine decides to be very relevant and do the same thing without informing the user, the problem of the extra time necessary for this operation could be a reason for the user to quit the dialog. The principle of transparency must be scrupulously respected : if the user knows what the machine is doing, length wait will be accepted more easily.

4. Statistical time evaluation of the system

A very important issue in the evaluation of a human-machine dialogue system available on the Web, is the time needed to reply to a user's sentence. In our system, there are four steps between the moment the user enters his statement in the dialog box and the moment he can hear and see the results. First, the sentence is sent through the Internet to the Xerox morphological tool. Then, the canonical form of this sentence is analysed by our concept recognition module in order to detect the important concepts to process. The third phase is a request to the INRIA's database accessible on the Web, and the last job of the system is to prepare a speech synthesis of the response and to send it to the Java applet client, with the hyperlinks and the textual information. In the following, we show some statistical analysis concerning the time needed for each module and the necessary time to provide a spoken interaction on the Web. We are not talking about the time needed for the speech recognition, because, at the moment, the module we are using is directly installed on the client machine, and connected to an IBM ViaVoice application; it is not therefore, possible to catch all these events from our server.

Time	Global	Xerox	% Xerox	Inria	% Inria	Elan	% Elan
Average	20.17	4.57	25.23	6.17	25.25	9.17	47.79
Standard deviation	6.12	0.67	7.83	3.40	14.40	4.94	12.23
Min	14.00	4.00	13.79	1.00	6.25	7.00	30.43
Max	32.00	6.00	37.50	11.00	47.83	19.00	65.52

Table 1 : Time statistics analysis of a man-machine interaction on the Web with the Halpin system

In our corpus, as we can see on Table 1 the average time needed for an entire interaction on the Web, from the moment the user validates his sentence till the moment he can hear the results, is 20.17 seconds. The morphological analysis represents 26 % of this time. The concept recognition is very quick (less than a ½ second and not mentioned on the figure). The access to the digital library is also 26 % of the time, and finally, the speech synthesis, with almost the half of the time, is the greediest module. In the case where there is no access to the database, the average time for a complete interaction is 12.36 seconds. It's about 45 % for the Xerox request and 55 % for the preparation of the speech.

5. Conclusion and future work

The HALPIN system is currently used by many people on the Web. The first results show that the users readily co-operate with the machine. This kind of multimodal natural language interaction is a valid answer to problems such as confusion, cognitive overload, and evaluation of the answer's relevancy. With the relative large number of real dialogs gathered on the Web by our system (more than 1000 files), we have a powerful tool for the study and development of a human-machine multimodal interaction model. We have shown that it is possible to dispatch through the Web a real time calculated speech synthesis, and we are now working on the integration of a voice recognition server module in our system. Some possibilities are currently being tested, such as the French "Janus system" [SCHULTZ 97] [AKBAR 98], in order to allow the user a free dialogue with the machine, in a more natural and effective way. We also count on the insertion of a large thesaurus, such as the French "Sémiographe" thesaurus from Memodata⁶, for a broader cover of the vocabulary in input as well as output.

[6] <http://www.memodata.com/>

6. References

- [AKBAR 98] Akbar M., Caelen J. (1998). Parole et traduction automatique : le module de reconnaissance RAPHAEL, *COLLING-ACL'98*, Montreal (Quebec), 36-40.
- [BATEMAN, HAGEN & STEIN 95] Bateman, J. A.; Hagen, E., & Stein A. (1995). Dialogue modeling for speech generation in multimodal information systems, in P. Dalsgaard, et al. (Ed.), *Proceedings of the ESCA Workshop on Spoken Dialogue Systems - Theories and Applications*. Aalborg, Denmark: ESCA/Aalborg University, 225-228.
- [BRUN 98] Brun, C. (1998). A Terminology Finite-State Preprocessing for Computational LFG. *36th International meeting of the Association for Computational Linguistics & 17th International Conference on Computational Linguistics*, Montreal, Quebec, Canada.
- [CHEYER & JULIA 95] Cheyer, A., and Julia, L. (1995). Multimodal maps : An agent-based approach in *Proc. of the International Conference on Cooperative Multimodal Communication (CMC/95)*, Eindhoven, The Netherlands.
- [COHEN 92] Cohen, P. (1992) The role of natural language in a multimodal interface, *Proceedings of the UIST'92*, 143-149.
- [CONKLIN 87] Conklin, J. (1987). Hypertext: an introduction and survey, *IEEE Computer*, 17-41.
- [GAUSSIER & al. 97] Gaussier, E., Grefenstette, G., Schulze, M. (1997). Traitement du langage naturel et recherche d'informations : quelques expériences sur le français. *Premières Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue de l'AUPELF-UREF*, Avignon, France.
- [GRICE 75] Grice H.P. (1975). Logic and conversation, in Cole, P., and Morgan, J.L., *Syntax and Semantics, volume 3, Speech Acts*. Academic Press, 41-58.
- [HARDIE 96] Hardie, E. (1996). A grain of sand or the ocean ; User aims in search engine interactions. *Fifth International WWW Conference - Poster Proceedings*, INRIA/CNIT, Paris La Défense.
- [PITKOW 96] Pitkow, J.E., and Kehoe, C.M. (1996). Emerging trends in the WWW user population. *Communications of the ACM* 39,6, 106-108.
- [ROUILLARD 99] Rouillard, J. (1999). Les enjeux d'un dialogue Homme-Machine sur Internet - L'Hyperdialogue. *Bulletin d'informatique approfondie et applications, revue de l'université de Provence*, 52, France, 3-20.
- [ROUILLARD 98] Rouillard, J. (1998). Hyperdialogue Homme-Machine sur le World Wide Web : Le système HALPIN, *ERGO'IA 98*, Biarritz, France, 343-345.
- [ROUILLARD & CAELEN 98] Rouillard, J. et Caelen, J. (1998). Etude du dialogue Homme-Machine en langue naturelle sur le Web pour une recherche documentaire, *Deuxième Colloque International sur l'Apprentissage Personne-Système, CAPS'98*, Caen, France, 99-119.
- [ROUILLARD 97a] Rouillard, J. et Caelen, J. (1997). Étude de la propagation au sein du Web à travers les liens hypertextes. *Quatrième conférence Internationale Hypertextes & Hypermédias*. Numéro spécial de la revue Hypertextes et Hypermédias, éditions Hermès, 1997, Paris, 87-100.
- [ROUILLARD 97b] Rouillard, J. et Caelen, J. (1997). A multimodal browser to navigate and search information on the Web. *Fourteenth International Conference on Speech Processing (ICSP97)*, IEEE Korea Council, IEEE Korea signal processing society, Séoul, Korea.
- [SCHULTZ 97] Schultz, T., Westphal, M., Waibel A. (1997). The GlobalPhone Project: Multilingual LVCSR with JANUS-3, *Multilingual Information Retrieval Dialogs: 2nd SQEL Workshop*, Plzen, Czech Republic, 20-27.
- [STEIN & MAIER 95] Stein, A. & Maier, E. (1995). *Structuring collaborative information-seeking dialogues*, Knowledge-Based Systems, 8 (2-3, *Special Issue on Human-Computer Collaboration*), 82-93.
- [STEIN & al. 97] Stein, A., Gulla, J. A., Müller, A. & Thiel, U. (1997). Conversational interaction for semantic access to multimedia information, in M.T. Maybury (Ed.), *Intelligent Multimedia Information Retrieval*. Menlo Park, CA: AAAI/The MIT Press, 399-421.