

# SEMANTIC MATCHING BETWEEN JOB OFFERS AND JOB SEARCH REQUESTS

COLING 90

José VEGA  
GSI-ERLI

72, Quai des Carrières  
94220 CHARENTON (FRANCE)  
Phone : (1)48 93 81 21, Fax : (1)43 75 79 79

The members of the development team were : B Carden, A Chaouachi, B Euzenat, G Klintzing, M Macary, R Leborgne. We wish to thank the LE MONDE newspaper team for their collaboration during the specification phase.

The primary objective of this system<sup>1</sup> which was developed for the LE MONDE daily newspaper, is to offer an efficient tool for a rapid and intelligent job searching service to professionals in the context of the ever increasing number of advertisements in the printed press.

Traditionally, offers of employment appear in newspapers and magazines and sometimes cover twenty pages. The person in search of employment faces the daunting task of daily readings of lists of job offers.

## FUNCTIONS PROPOSED BY THE SYSTEM

This system carries out an optimized comparison between the job offers in the advertisement data base and the requests and/or curriculum vitae entered by end-users at their terminals (minitel). This is performed by extracting pertinent information from the input texts and comparing it at semantic level with the Knowledge Base and the data of the indexed offers.

The requirements expressed in the job offer should resemble as closely as possible the characteristics of the candidate or at least be semantically close. More precisely, the results of the matching process are grouped according to the three following criteria :

- the requirements of the position are directly fulfilled by the characteristics of the candidate
- the requirements expressed in the job offer are met only partially
- the job offers require characteristics other than those expressed in the candidate's curriculum vitae, but are in the same semantic field.

For example, if the requested position is:

(1) *'project manager real time computing'*,

<sup>1</sup> This system is in operation since September 89 and can be consulted by Minitel, using telephone number 3615 and selecting the LM / EMPLOI service.

the system will propose to the candidate 3 categories of job offers :

1.- *Project manager real time*

- *Project manager in process control*

- *Project manager in automation & industrial computing*

- *Head of process control department*

-----  
2.- *Project manager in software engineering*

-----  
3.- *Project manager in computing*

## ARCHITECTURE

### The End-User Interface

The end-user interface allows the candidate to enter his curriculum vitae in natural language<sup>2</sup>.

Part of the interaction with the candidate concerns the identification of unknown words (typing errors and spelling mistakes) at which point the system asks the user to correct his text. When a user's request does not match with any requirements in the job offers data base, the system enters into a dialogue with the candidate in order to relax or modify the constraints he imposed on the job search criteria at initial request time.

The job offers proposed are assembled into groups according to their semantic pertinence with respect to the candidate's request.

### Databases

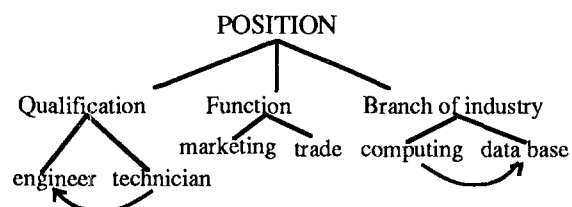
The system uses linguistic data :

- 
- 2- The user can express his attributes freely and without constraints, contrary to SQL, for example.

- a general dictionary containing grammatical words, verbs and a certain number of nouns;
- a dictionary specialized in the universe of employment (professions, training, universities, software tools, regions,...);
- a Knowledge Base (KB)<sup>3</sup>.

The conceptual, semantic and pragmatic models of this application are represented in the KB. This KB describes certain facts which are universal truths and others which are only true within the context of the universe of employment.

The job offers and the curriculum vitae of the candidates (or users) have been modelled using the object analogy. Each conceptual object has an associated attribute list (domains) with values which instantiate them. The values of a particular domain are linked together by semantic and pragmatic relations. In the same way, relations may exist between values of different domains. For example :



where POSITION is an object. Qualification, Function, ... are the domains. Marketing, computing, ... the values. Moreover, we can see an "upper-level" relation between technician & engineer, and a generic term relation between computing and data base.

### The analysers

The system uses morphologic and syntactic analysers and a semantic analysis engine called the "matching machine" (MM).

The rule sets (see below for further details) and grammars used by the morphologic and syntactic analysers in this application were designed to be linguistically robust and rapid in execution. Given that the application was designed for 200 simultaneous Minitel<sup>4</sup> connexions by the response time for these analysers must be extremely short.

- 
- 3- The combined size of the two dictionaries is approximately 30,000 words with 3000 referential words for the KB.
  - 4- This physical architecture consists of a : frontend which manages the connexions and serializes the user's queries, and a backend supporting the analysers.

With regard to questions of morphology, the analysers possess rule sets describing inflexion and derivation for the recovery of canonical forms of words<sup>5</sup> stocked in the dictionary starting from the text of the user's request or curriculum vitae.

This analyser also possesses rules for treating initial letters (H.E.C. <==> HEC, CIA <==> C.I.A., ...), abbreviations (Sté <==> Société, m <==> mètre, ...), "floating prefix" terms (micro-informatique <==> microinformatique <==> micro informatique), concatenated or disjoint expressions (mettre en oeuvre <==> les mesures mises rapidement en oeuvre ..., pomme de terre, ...) and other morpho-lexical phenomena.

Concerning syntactic analysis, the corresponding analyser possesses a grammar of "standard" French. However, phenomena such as anaphora, coreferencing (except in certain minor cases), the scope of negations, among others are not treated. This is a deliberate choice since the persons using this system (through their requests or curriculum vitae) do not often use these elements of style in their texts (texts are chiefly noun phrases or verbal sentences).

It is important to note that the analysers described above<sup>6</sup> are independent of the application and can be reused for other applications.

Concerning the text comprehension phase, the MM treats the information received from the syntactic analyser in conjunction with information drawn from the Knowledge Base.

The MM uses functions or "methods" which carry out specific treatments according to the type of objects under consideration.

### How the Matching Machine works

The functioning of the MM is at the same time semantic and pragmatic and 4 distinct steps are identified. They are :

- 1. Recuperation of normalized terms from the user's request or curriculum vitae;
- 2. Identification of the domain and of the object concerned by these terms;
- 3. Semantico-pragmatic spreading from the initial terms according to the "method" used for their associated object.

- 
- 5- For us, canonical words are : a singular, masculine nouns or adjectives, and roots of verbs.
  - 6- except some rules used to handle special words like the acronyms, the "telematic language", etc.

- 4. Extraction, intersection and classification of the indexed job offers according to the initial terms and those identified by the spreading process.

Step 2 serves to unambiguously identify the objects designated by the normalized terms which were extracted from the user's request. For example, in the following request:

(2) *Expert translator of text in English*

the analyser will assign the term "English" to the domain FUNCTION of the object POSITION since this term designates, in this context, a specialization within the profession of the translator.

In contrast, if the request is:

(3) *Civil engineer speaking English*

in this case, the term "English" will be considered, in this context, as a value designating an object LANGUAGE (which is one of the conceptual level objects found in the job offers)<sup>7</sup>.

Step 3 consists in passing from one term to another, starting at an initial term, in a tree-walk through the semantic and pragmatic network of the KB. This is performed in an outward spreading manner and is determined by the methods associated with the object types designated by the initial terms. The arcs between the nodes of the network are weighted and the result of a spreading process is a new term Y at a distance n from a starting term X.

The distance that a spreading process is allowed to run through the network is determined by the methods. This distance is one of the parameters necessary to calculate the final distance in the following step.

Step 4 is charged with the ordering of the job offers by comparing initial and final terms.

The sets of job offers then undergo set operations (boolean operations). This treatment is directed by a number of dynamically acting rules. That is, the actions of these set operations depend on the semantic role assigned to the terms during the second step of analysis and the objects concerned by these terms.

For example, for the request:

(4) *Computing journalist*

---

<sup>7</sup>- Among the examples mentioned here, we could consider the following:

(6) *English translator*

Given that the model does not take nationalities into account, the system will interpret this request as in example (2).

the job offers proposed must correspond to positions for journalists specializing in the computing domain and not to positions in the press and/or informatics.

However, in the following example:

(5) *UNIX / C programmer*

the system must propose positions for specialists in UNIX / C. It will also propose job offers for software programmers in which no mention of operating systems or programming language is made, and others in which other operating systems or languages were mentioned.

The classification of job offers is made as a function of the distance and the criteria fulfilled by the request. The job offers will be presented to the user according to this classification (see Example (1)).

## CONCLUSION

Although we use natural language processing as a communication interface, we have not neglected the strong points of 'formatted-screen based dialogue'. Indeed, much emphasis was given to the design phase of the ergonomics (screen content, fields, messages...) and cinematics (dynamic of screens as a function of the users' actions) of the application in order to integrate the better aspects of both dialogue modes into the interface.

In fact, the experiments we conducted during the specification phase demonstrated that, in such an application, it is not feasible to present to the user a virtually blank screen containing brief suggestions such as "Enter your CV" or "Enter a request".

As a result we designed the interface to present to the user fields corresponding to conceptual objects in the knowledge base in which he is allowed to express himself without lexical or syntactic restriction. In this way the user feels guided through the session without being constrained to reply to imposed questions.

By integrating natural language processing into the application we have implemented a new approach to man-machine dialogue in the field of online job offer services (contrary to the classical arborescent menu-driven interfaces).

The choice of combining the advantages of formatted-screen based dialogue and the constraint-free approach of natural language text input makes a significant advance towards the design of effective user-friendly man-machine interfaces.