THE NEED FOR INFORMATION AND KNOWLEDGE MANAGEMENT: an integrated approach for the Quebec Civil Service

Suzanne BERTRAND-GASTALDY
Ecole de bibliothéconomie et des sciences de l'information
Université de Montréal
P.O. Box 6128, Station A
Montreal, QC
CANADA H3C 3J7
GASTALDY UMTLVR(Bitnet)

Luc DUPUY & Louis-Claude PAQUIN
Centre d'analyse de textes par ordinateur (ATO)
Université du Québec à Montréal
P.O. Box 8888, Station A
Montreal, QC
CANADA H3C 3P8
R21014 UQAM(Bitnet)

Proposal for the Second International Congress - Terminology and Knowledge Engineering Applications, Section 7: Information Management in Organizations

I THE CONTEXT

The production and use of documents in the Quebec Civil Service

The Quebec Civil Service (i.e. government office workers who are employed in thirty ministries and over two hundred affiliated agencies) generates and uses a great variety and body of texts: legal documents (the laws of the province, judicial decisions, tenets, regulations, decrees, norms, rulings, rules, tariffs, letters patent, collective agreements, employee grievances and their settlement), administrative documents (policy statements and directives), technical documents (reports of many kinds) and print and electronic correspondence. To these we must add, for they too are texts in their own right, questions sent by civil servants to the administrative units concerned and those also of the public, as well as their replies.

Not only is the literature voluminous (the administrative policy alone consists of several hundred items, and there are over 520 laws, over 2000 rulings, 110 collective agreements, etc.) but it grows and is modified on a daily basis. For example, the Sécurité du Revenu creates an average of five new documents a day, 6000 requests of dispensation are sent to the Greffe each year, and the Recueil de jurisprudence must account for the quarter million decisions reached annually by the different courts of Quebec. There are, literally, information bottlenecks and the circulation of printed updates cannot keep up with the accelerated pace of production: a modification in administrative policy promulgated late last May and available the next day in the experimental textual database of the Secrétariat du Conseil du Trésor, came out in print in mid-October. Since all documents are closely linked, a modification in one can have repercussions on several hundred others. Changes in the civil code for example could affect 237 laws. Inserting these in ring binders entails considerable expense for each administrative unit.

The consequences

It is quite clear that under these conditions no government worker can be cognizant of all the documents nor all their modifications: the Direction des évaluations gouvernementales of the Ministère de l'Environnement now has at its disposal thirty 200-page impact studies, thirty-two 150-page volumes of public hearing reports, about a hundred thematic studies and sixteen legislative rulings, as well as the day-to-day production of its office. Consulting these documents is difficult for several reasons other than their volume and heterogeneity.

Indeed, printed texts are stored in bulky binders that are hard to manipulate. Cross-references to other documents are many and these documents are not always available.

Moreover, human indexing costs are so great users must content themselves with reduced access through tables of contents and summary indexes that give few relations between synonyms, hyperonyms and hyponyms: for example, the phrase "frais de voyage" [travel expenses] used by government workers cannot be located because administrative policy only mentions "frais de déplacement" [transportation expenses]. Indexing more thoroughly with a partially controlled vocabulary only causes penalizing delays. For instance, the index of the Refonte des lois provinciales isn't published to this day yet the document dates back to 1977. Since the legal documentation is abundant and scattered about, it is almost impossible to know the actual state of the legislation following any changes, modifications, abrogations or substitutions it may have undergone.

It is then understandable that the government worker pressed for a decision or an answer is disheartened at the very thought of searching through this mass of documents. What came out in an informal study conducted during the planning stages of an expert system designed to assist in the application of one area of administrative policy, is that the worker will resort to the usual bag of tricks if print information is not easily accessible: he will rely on hiscontacts and ring up colleagues who may have had to deal with the same problem.

Information habits vary from one individual to the next: "youngsters" somehow manage at first without knowing all their way around and with experience, these "oldsters" develop time-saving strategies that mostly short-circuit official sources but the actions taken may be ill-founded. In order to avoid this, controllers are hired, validation is further increased and delays and costs multiply. At the same time, resources are diverted that otherwise would be invested toward improving access to authoritative documents. The situation in the end is paradoxical for those who must enforce policies are not even aware of the texts of these policies!

II THE USUAL SOLUTIONS

Information retrieval and expert systems

Two separate entities

Generally speaking, solving information problems in the private or public sector consists in handing over the key to one of the following systems: an information retrieval system that gives access to documents in which the user searches himself for the information he needs, or an expert system that supplies answers without allowing the user to validate them by checking the texts. We should also add here external bibliographic database query systems and decision support systems that pertain for the most part to numerical data.

According then to the nature of the problem at hand, the user will address one or the other system, without common interfacing, without knowing if they might intersect or not or if on the other hand they might not answer all needs, and without knowing if the information retrieval and expert systems rely on the same sources.

Two entities designed without a global view of actual user problems

If several solutions are available however, it isn't because the user's predicament requires them but because they are implemented by different specialists who do not actually work together and only seesome aspects of the problem. They will of course assess user needs according to their own methodology, tools and points of view. In fact, the final proposal is more often influenced by the latest technology than by actual user needs. The cooperation advocated between information specialists and "knowledge engineers" (Morris & O'Neill, 1988, p. 176), has yet to become common practice.

Two entities designed without involving users

A frequent complaint touches upon the poor usefulness, despite their number, of user studies, the inaccuracy of their results and the lack of success of the systems in place. It is believed that users have difficulty expressing their needs and it isn't quite known how they use information (Brittain, 1987, p. 150). But are their needs and "corporate" culture really taken into account? They do form within the organization a micro-society of their own, with its codes, its jargon, its patterns of reference, its particular structuring of the world, its power plays, etc.

Two entities designed to occult the texts

Each system also has the effect of occulting the texts depositary of the structures of representation of organizational culture.

Bibliographic information systems depict the original texts by a series of juxtaposed key-words and in the best of cases, by an "informative" abstract that are as many interpretations on the part of an admittedly "intelligent" indexer who cannot help but be subjective and influenced by his apriorisms, his preoccupations at that time and his own knowledge structure. Such indexing sets only one "reading" of the text whereas each subsequent reading, were they allowed, would lead to different interpretations. When indexing is carried out by retrieving terms automatically in all complete texts, it is faster, more uniform ("coherent") and impartial but only covers surface forms (character sequences) and neglects the many underlying structures that contribute to the meaning of the text.

If a thesaurus is developed to represent concepts beyond the various formulations that may describe them, it is most often constructed a priori with other terminological tools by a select few "specialists". Too seldom do they turn to the different contexts of the terms in the corpus and then

proceed with their conceptual organization. From the onset, tension builds between the concepts of the text and those of the thesaurus. What will happen once the thesaurus is updated?

Lee (1985) notes that there is a greater need for extensional semantics in organizations than for intentional semantics. The concept of student for example can be apprehended in relation to how the Quebec Civil Service employs students, that is on a temporary basis in the summertime. This is what comes through in the definition given in the administrative policy of the Conseil du Trésor. The definition given by the Ministère de l'Education is another matter entirely. An information system designed for the Quebec Civil Service must then take into consideration the semantics of those many micro-worlds that are the ministries and their divisions. The several context uses of terms must also be summoned and the system semantics must reflect any and all modifications that come up in the texts.

These micro-worlds all interact to create a larger culture that isn't the mere sum of its parts but the result of negotiations. A meta-thesaurus should be constituted to take into account interrelations between local thesauri.

The same phenomenon is observed in constituting expert systems. Indeed, most knowledge engineering methodologies (the retrieval and representation of knowledge in expert systems) are inspireds from cognitive psychology: protocol analysis, personal construct theory, etc. These approaches tackle knowledge from without with no reference to context, neglecting as it were an important source of data which experts store themselves in textual archives: their articles, educational papers, etc. We believe this overshadows the social dimension of expertise.

In public and para-public organizations, expertise lies in the interpretation of a corpus of prescriptive texts and reveals potentially divergent if not downright contradictory aspects. A notion such as "l'intention de frauder" [the intent to fraud] for example, is difficult to ascertain ab abstracto and turning to the contexts supplied by the body of archives (the laws, decisions and rulings) is a necessary step toward its definition.

In other respects, uncovering the knowledge of an expert means reconstituting the context in which the coalescence of scattered information, the learning process, took place. If the implicit part of knowledge is to be significantly reduced, the reconstruction must then be made according to an explicit discursive process, by voice (the interview) or in writing (the handbook).

Last, we must bear in mind that any form of knowledge depends in part or in whole on a socio-linguistic structure - morphologies - that allows the stocking, handling and transmission of elements of a particular field of knowledge. Concept screening based on morpho-syntactic patterns

increases exhaustivity, the scope of sources used and terminological precision. The search is not conducted from what may be envisioned from the data but from the data itself, or from analogies drawn with other fields. Experts should only be consulted afterwards to validate and trim the material gathered.

III THE SOLUTIONS ADVOCATED BY THE QUEBEC CIVIL SERVICE

In order to counter the disastrous effects of the solutions proposed until now, the Quebec Civil Service adopted a multidisciplinary approach centered both on the texts and their users.

Following several studies and projects in the field of information management and expert systems, a group was formed bringing together researchers who had closely worked with government workers in carrying out expertise transfer evaluations and conceiving information systems in several workplaces.[1] The group also welcomes workers from three test sites.

Computer science intervention in information matters seen as a continuum

The literature suggests at times that information retrieval and expert systems are but two extremes of one continuum and that if no intermediate solutions are recommended, it isn't for lack of correspondence with user

[1]-The group consists in a first instance of François Daoust, author of the SATO program (Système d'analyse de textes par ordinateur), Louis-Claude Paquin, author of the D-expert program, and Luc Dupuy, all three members of the ITC section (Ingénierie textuelle et cognitive) of the Centre d'analyse de textes par ordinateur (ATO) at the Université du Québec à Montréal, and experienced in the implementation of textual knowledge-based expert systems. The two other members of the group hail from the Ecole de bibliothéconomie et des sciences de l'information at the Université de Montréal, and are specialized in information systems and languages: they are Gracia Pagola, research assistant, and Suzanne Bertrand-Gastaldy, who presides over the group.

needs but because specialists are still hard pressed to work together (Morris & O'Neill, 1988, p. 178). Users are the unwitting victims of this situation. Salton (1985) foresees a unified approach of future information systems (databases, bibliographic systems, question & answer systems) and some researchers have already drawn a structure integrating databases, information systems and knowledge-based systems (Bell, 1985).

Our group strives for more than software integration. We have elected to work toward a methodological integration centered around the text, now the prime object around which the information retrieval and expert systems and the lexical databases revolve. We believe however there is quite a gap between computer science methodology and regular socio-analytical methods.

Accounting for the many facets of information

Our study is then based on a corpus of texts produced and used in the Quebec Civil Service. We apprehend these texts by quantitative and qualitative methodology with the help of the SATO program.

SATO is a tool box of sorts developed for text content analysis. It is a textual database system that can constitute a corpus, outline and store structured lexical data, annotate and manipulate multilingual texts. SATO operations consist in locating textual segments, constructing lexicons, categorizing terms, drawing lists, lexico-syntactical analyses, etc.

The program allows the user to question his text and control each step of its treatment. To this end, SATO must first read the text and recognize its components: words, paragraphs, page markings, etc. The results of this preliminary treatment are stored away in files specifically designed for SATO. At this step, SATO also constructs a lexicon, a catalogue of all lexical items found in the text.

An interesting feature of the system is that it is capable of associating numerical or symbolic properties to terms in the texts. It is then quite possible to associate a grammatical or semantic category with a term and modify property values, without ever altering the text whose integral body is thus sustained.

SATO offers an array of resources that can be used as is or combined to others to develop other analyzers. The analytical strategies developed here can be stored for later use. The system has the dual advantage that it can tailor itself to respond to the hypotheses of experimented users or be programmed beforehand for beginners. Without going into detail, we can say that SATO allows textual explorations that are in no way comparable to the usual full text automatic indexing and retrieval programs.

The generation of lexicons, listing the number of occurrences of each form, can pertain to the entire text or to any one part of the text, paving the way for comparisons between corpi or different parts of the same text. Establishing a terminology listing specialized terms is done all the more faster by applying grammatical categorizer mechanisms and noun phrase screening procedures to texts. It is also possible to consult a dictionary already on file and apply it on a new corpus.

Posting concordances, and determining the context ("n" words, the sentence and paragraph, a passage indicative of a particular morphology such as the use of quotation marks, the phrase "on entend par", etc.), can also be useful in pinpointing definitions, writing and setting terminological files, locating homographs and polysems, and in searching for syntactic and semantic relations or predicate attributes. These are definite assets in the a posteriori elaboration of a thesaurus or the design of a knowledge database.

It is also possible using SATO to locate words or classes of words that distinguish the texts or sub-texts, a point of interest for the indexer who is most preoccupied by selecting key-words with great discriminating value. Various comparative analyses can be made on the texts, sub-texts and lexicons, highlighting significant regularities and irregularities.

SATO thus opens onto an information system that preserve the texts while enriching them with countless knowledge - that of each user and that of all users. It is a good solution in that nothing is set in stone and users can refer to it at will according to their own needs or reading hypotheses.

Taking into account the very real problems faced by users

The study of textual productions follows one of user needs that consider foremost their comments on how they perceive and accomplish their work: this approach harks back to the INSTRAT Project which strove to find solutions from the nature of the problems at hand (Belkin, Seeger & Wersig, 1983). An inventory can then be made of interactions users have or should have with knowledge and information sources: having select access to textual content, reading and providing needed answers; synthetizing the knowledge set in the texts and writing up an expert opinion or rewording texts to make them more coherent; consulting an expert to carry out a limited task that falls without the range of competence, such as filling out an expense report for example; questioning external databasesto keep up to

date, etc. This project has developed to the point where it is now financed.[2] Each task requires the assistance of complementary systems - or modules for that matter - that must be interconnected.

Training users to take charge of their information resources

The study of user needs continues throughout the project.

Several animation and training sessions bring together the researchers and the workers involved in the test sites. From the tasks to be accomplished to real work situations, they are shown the means to solve their problems, the underlying theoretical notions are explained to them, to the extent that they are now more capable of specifying their needs. We do not simply teach them about the functions of the program, we "equip" them with concepts that will enable them to use computer-assisted reading modes to index, synthetize or modify their texts. They are the ones who know what they want done on the texts; the analysts simply follow their lead. It is in such a spirit that users are integrated in the system development process, thus ensuring later and better use of these systems. We advocate that the validation and control of analytical operations be left to the users, limiting external consultations to animation and the methodological training of users.

This corresponds to a trend observed by Haselkorn (1988, p. 4): "[...] the computer industry is realizing that while engineering advances are essential, they must be tied to helping people do something they want to do". If the poor quality of full-text information program functions found on the market is any sign, software designers should jump on this wagon, the sooner the better.

CONCLUSION

In short, the study of needs that we promote subjects the elaboration of an information system for the Quebec Civil Service to the texts as well as to user idiosyncrasies. The system implemented will have to be polyvalent and integrated in order to properly answer all the needs of these "text workers",

---

[2]-F. Daoust, L. Dupuy & L.-C. Paquin, "ACTE: a workbench for knowledge engineering and textual data analysis in the social sciences", ICEBOL4: Fourth International Conference on Symbolic and Logical Computing, 1989, forthcoming.

those for whom reading and analyzing texts constitute major activities. They cannot settle for a partial solution to their information needs, nor can they settle for one not tailored to their organizational culture.

We believe that the success and steady, efficient use of such a system depends on the involvement of users in the planning stages. This requisite involvement partakes of an effort to make known the usefulness of software tools and text analysis methodology, and their impact on performance as to speed and control on the task accomplished.

Apart from the information system itself, useful by-products are generated for the organization: the constitution of textual databases and the implementation of a protocol for their rational management, and the compilation of lexical databases that standardize and structure the terminology of the concepts handled.

This research/action approach makes the organization aware of the wealth of knowledge found in its texts while allowing it at the same time to investigate its history to further elaborate its present stands. It also throws new light on the conflict between those who enforce rules and regulations and those who hold the explanation by increasing the instances where workers voice their opinions on their "organizational culture".

Translated by Dominique Michaud