

Generic dialogue modeling for multi-application dialogue systems

Trung H. Bui, Job Zwiers, Anton Nijholt, and Mannes Poel

Human Media Interaction, Department of Computer Science, University of Twente
Postbox 217,
7500 AE Enschede, The Netherlands
{buih, zwiers, anijholt, mpoel}@cs.utwente.nl

Abstract. We present a novel approach to developing interfaces for multi-application dialogue systems. The targeted interfaces allow transparent switching between a large number of applications within one system. The approach, based on the Rapid Dialogue Prototyping Methodology (RDPM) and the Vector Space model techniques from Information Retrieval, is composed of three main steps: (1) producing finalized dialogue models for applications using the RDPM, (2) designing an application interaction hierarchy, and (3) navigating between the applications based on the user's application of interest.

1 Introduction

A multi-application dialogue system is defined as a dialogue system allowing the user to navigate between a set of applications. Applications considered range from simple tasks such as operating a home device or booking a flight to more complex tasks such as intelligent traffic management or smart-room control in the crisis management domain.

To date, due to the complexity of the management of language interfaces and their strong dependence on the interaction context, a really generic approach for dialogue design does not yet exist; each application requires the development of a specific model. Dialogue model prototyping therefore represents a significant part in the development process of interactive systems. However, most current design methods are limited to the development of dialogue systems working on a single application or a small set of applications [3], [11], [9], [12].

In this perspective, we aim to develop a generic dialogue modeling methodology for the efficient production of interfaces for multi-application dialogue systems (Fig. 1). The targeted interface allows transparent switching between a large number of applications within one system. The approach, based on the Rapid Dialogue Prototyping Methodology (RDPM) [1] and the Vector Space (VS) model techniques from Information Retrieval, is composed of three main steps: (1) producing finalized dialogue models for applications using the RDPM, (2) designing an application interaction hierarchy, and (3) navigating between the applications based on the user's application of interest.

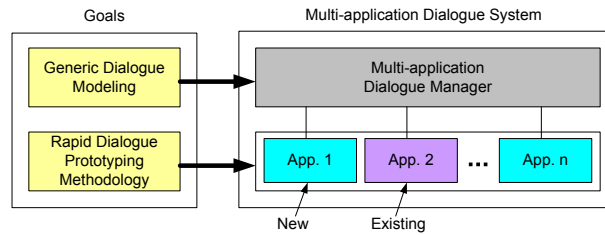


Fig. 1. Multi-application dialogue system development.

These steps are described in sections 2, 3, 4 respectively. A scenario example for producing a dialogue system accessing 10 applications in the ICIS domain¹ is presented in section 5. Finally, in sections 6, 7 we summarize the main points of the paper and possible further extensions of the methodology.

2 Producing finalized dialogue models for applications using the RDPM

The finalized dialogue model for each application (cf. Fig. 1) can be quickly produced using the RDPM. The general idea underlying the RDPM is that the dialogue model is a frame-based model that can be quite easily and systematically derived from a relational representation of the application itself, hereafter called the task model. More precisely, the RDPM consists of five main consecutive steps, namely: (1) *producing a task model* for the targeted application; (2) automatically *deriving an initial dialogue model* (and the associated multimodal dialogue-driven interface) from the produced task model; (3) using the generated interface to *carry out Wizard-of-Oz experiments* (i.e. dialogue simulations) to improve the initial dialogue model; (4) *carrying out an internal field test* to further refine the dialogue model (reformulation of system messages, improved feedback, etc.), and to validate the evaluation procedure (coherence, understandability); and (5) *carrying out an external field test* to evaluate the final dialogue model according to the evaluation procedure defined during the internal field test. Steps 1, 2 are briefly described in the next sections in the context of producing finalized dialogue models for applications, the remaining steps are described in detail in [1].

2.1 Task model

In the RDPM, an application is seen as a set of functions the user can invoke through the multimodal interface to perform the various functionalities provided by the application. In this perspective, an application is modeled as a set of relational tables, where the rows correspond to the possible functions (also called

¹ <http://www.decis.nl/html/icis.html>

“solutions” or “targets”) and the columns are the attributes needed to uniquely identify each of the functions, and to invoke it. In other words, the values of the attributes in a row of the solution table (also referred to as canonical values) correspond to the values of the arguments of the function, the call of which results in the fulfillment of the corresponding application functionality. For example, in the ICIS domain, the task model for the patient search can reduce to a single generic function `select_patient(name, age, address, ...)`, the attributes of which identify the selection features available for the patient search. Therefore, the task model of the patient search is simply a table with as many columns as there are attributes, the rows of which are the various value combinations corresponding to patients. At the computational level, the calls to the `select_patient()` function are implemented in the form of SQL queries to the project database containing the required information.

2.2 Finalized dialogue model for a single application

In our approach, a finalized dialogue model is defined as a set of interconnected multimodal Generic Dialogue Nodes (referred to as mGDNs [8]), where each of the dialogue nodes is associated with one of the attributes (also called “slots” or “fields”) in the solution table. In complex applications, these mGDNs are divided into groups, where each group element is considered as an object and the mGDNs in the group are attributes of the object. For instance, the *First Name*, *Last Name*, and *Function* mGDNs belong to the *Person* group. For any given slot, the role of the associated mGDN is to perform the simple interaction with the user that is required to obtain a valid value for the associated attribute. In the architecture that we have selected for the implementation of our multimodal dialogue-driven interfaces, the processing of the mGDNs (i.e. the actual interaction with the user according to the specification of the mGDNs) is performed by a specific module called the *local dialogue manager*. However, this is, of course, not sufficient to carry out any real dialogue: some form of global dialogue management also has to be integrated. For example, in addition to the definition of the mGDNs and the specification of the local dialogue manager, some branching logic responsible for the management of the global dialogue flow needs to be specified. In our approach, this branching logic is hard-coded in a specific dialogue management module, called the *global dialogue manager*. The underlying assumption is that the encoded local and global dialogue flow management strategies are indeed application-independent, i.e. that, in most situations, they lead to an acceptable, though not always optimal behavior for the system. Consequently, in our approach, dialogue model design essentially reduces to the application-dependent, declarative specification of the mGDNs, the encoded dialogue management strategies being used without modification for all applications. In short, a finalized dialogue model consists of two main parts: (1) the application-dependent, declarative specification of the mGDNs; and (2) the application-independent (local and global) dialogue flow management strategies encoded in the corresponding (local and global) dialogue manager. Both of these components are described in more detail in [1].

2.3 System Architecture

The general architecture of the dialogue system corresponding to each single application produced by the RDPM is represented in Fig. 2.

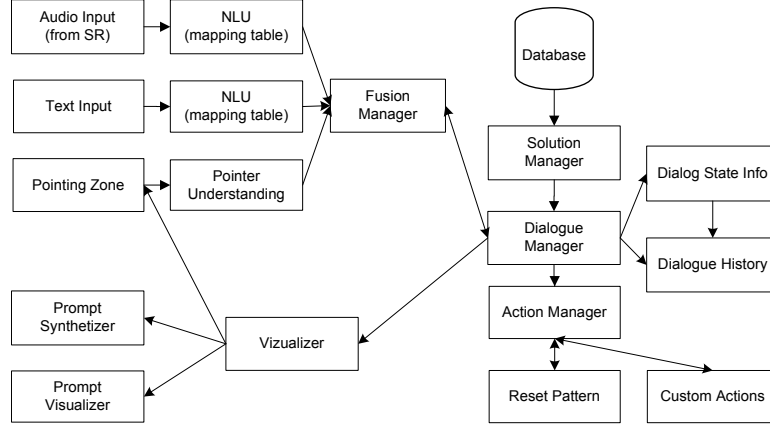


Fig. 2. Architecture of dialogue systems produced by RDPM.

Three input modalities: voice, text and pointing can be used independently or simultaneously depending on the configuration of the current active mGDN [8]. These inputs are pre-processed by the Natural Language Understanding (NLU) modules and the Pointer Understanding (PU) module before sending to the fusion manager. The fusion manager integrates the semantic triples (attribute, value, time-stamp) receiving from the NLU and PU modules and sends the output semantic triples to the dialogue manager.

The dialogue manager encodes the local dialogue flow management strategy and global dialogue management strategy. Therefore, the input to the dialogue manager is first processed by the local dialogue management strategy in which we define five types of generic situations: *OK*, *Request for Repetition*, *Request for Help*, *NoInput*, and *NoMatch* [1].

In the case of the *OK* situation, control is handed back to the global dialogue manager which applies the global dialogue management strategy for the activation of the next mGDN. When the dialogue manager gathers enough constraints², it sends the request to the action manager, the application connected with this module performs the task and sends the feedbacks to the action manager, the action manager then forwards these feedbacks to the dialogue manager. In addition, functions related with user modeling and system customization have

² This happens when the number of solutions (extracting from the solution manger) satisfying the current constraints is smaller than or equal to a pre-defined solution threshold.

been integrated. In particular: (1) *Reset Patterns* that allow the system to adapt to the behavior of a specific user or population of users by anticipating their next decisions, and (2) *Custom Actions* that allow the users to dynamically associate sequences of solutions with a single new solution. The main goal of these extensions is to reduce the time to perform a task with the interface. The hypothesis is that these functions will indeed increase the quality of the interaction as perceived by the user. These two functions are described in more detail in [2].

The output of the dialogue manager module to the visualizer module are multimedia prompts containing messages and a pointing zone. The messages are visualized in the user interface (Prompt Visualizer) and/or uttered by the mGDN during the interaction (Prompt Synthesizer). The messages are combined with a pointing zone (the content of this zone is a map, a calendar or a table depending on the nature of the mGDN) to allow the user to provide the desired values using keyboard, microphone or mouse click/touchscreen.

3 Designing an application interaction hierarchy

In section 2, using the RDPM it is possible to produce n finalized dialogue models M_0, M_1, \dots, M_{n-1} from n applications A_0, A_1, \dots, A_{n-1} , the question is how to integrate these applications in one unique system (i.e. multi-application dialogue system).

Vrugt and Portele [12] introduced a dialogue system accessing multiple applications with a dynamic setup that can be changed at run-time. Their goal is achieved by application-independent knowledge processing inside the dialogue system based on modular ontological descriptions. They also defined a clear interface between a dialogue system and applications by realizing a generic dialogue functionality on top of the application independent knowledge processing.

Carroll and Carpenter [3] developed a call-routing dialogue system using the VS model techniques from Information Retrieval. The system allows routing automatically the user's telephone call to the right department.

We organize applications in a hierarchy since it allows flexible dealing with a large number of applications [4]. The hierarchy can be created manually or automatically. When the number of application is large (hundreds, thousands, or more), it is difficult to create the hierarchy manually, therefore an automatical process is suitable for this case. In our approach, the hierarchy is produced automatically using VS model techniques and clustering algorithms.

3.1 Application interaction hierarchy

An application interaction hierarchy is an m levels hierarchy of n finalized dialogue models consisting of three types of nodes:

1. Root (level: $m-1$): unique node on the top of the hierarchy.
2. Internal nodes (level: from $m-2$ to 1): each internal node consists of at least two child nodes, a child node can be an internal node or a leaf node. The

hierarchy accepts lattice nodes (i.e. internal nodes, each of them has more than one father node).

3. Leaves(level: 0): correspond to n applications.

An application interaction hierarchy ($n = 10$) is represented in Fig. 5.

3.2 Vector space model for the finalized dialogue models

Each finalized dialogue model produced in section 2 is characterized by a textual description of the associated application. We represent these descriptions by k -dimension vectors d_0, d_1, \dots, d_{n-1} using the VS model techniques.

The following paragraph presents the process of producing vectors and computing the similarity between the applications using the standard VS model technique (in the implementation phase, a suitable VS model is selected based on the content of textual descriptions³):

1. Produce index terms from the textual descriptions

We analyze the textual descriptions using Natural Language Processing (NLP) techniques (syntactic analysis, morphological & stop words filtering, term extraction) to produce k index terms: t_1, t_2, \dots, t_k .

2. Construct occurrence matrix F

A description is represented by a lexical profile: $d_i = (w_{i0}, w_{i1}, \dots, w_{ik-1})$. w_{ij} is the weight (or importance) of the j^{th} indexing term t_j in the textual description d_i . w_{ij} is often simply the number of occurrences of t_j in d_i or the inverted occurrence frequency.

The $n \times k$ occurrence matrix F:

$$F = \begin{pmatrix} d_0 \\ d_1 \\ \dots \\ d_{n-1} \end{pmatrix} = \begin{pmatrix} w_{00} & w_{01} & \dots & w_{0k-1} \\ w_{10} & w_{11} & \dots & w_{1k-1} \\ \dots & \dots & \dots & \dots \\ w_{n-10} & w_{n-11} & \dots & w_{n-1k-1} \end{pmatrix}$$

3. Compute the score(or measure the similarity)

The most common similarity measure for the standard VS model is the cosine of the angle between the vectors:

$$sim(d_i, d_j) = \cos(\vec{d}_i, \vec{d}_j) = \frac{\sum_{p=0}^{k-1} (w_{ip} \times w_{jp})}{\sqrt{\sum_{p=0}^{k-1} w_{ip}^2 \times \sum_{p=0}^{k-1} w_{jp}^2}}$$

We use this measure to determine the similarity between two applications, i.e. the score between A_i and A_j : $s(A_i, A_j) = sim(d_i, d_j)$.

³ For example, in case the textual description is a set of sentences, a semantic VS model taking into account the dependence between terms such as [10] is appropriate.

3.3 Hierarchical clustering algorithm

From the vectors d_0, d_1, \dots, d_{n-1} and their similarity computation produced in 3.2, we use the hierarchical clustering algorithm [7] to produce the application interaction hierarchy:

1. Consider each d_i is a single cluster, we have n clusters, the distances between a pair of clusters i and j (in this step): $D(i, j) = 1 - sim(d_i, d_j)$.
2. Find the most similar pair of clusters (i.e. $\min(D(i, j))$) and merge them into a single cluster, so that we have one cluster less.
3. Compute distances between the new cluster and each of the old clusters.
4. Repeat steps 2 and 3 until all items are clustered into a single cluster of size n .

Step 3 can be done in several ways such as single-linkage, complete-linkage, or average-linkage clustering [6]. Applying the single-linkage, the formula to calculate the distance between two nodes i, j :

$$D(i, j) = \min[D(i_p, j_q)]$$

i_p, j_q are leaves belonging to i and j respectively, $p \in [0, n_1 - 1], q \in [0, n_2 - 1]$, n_1 and n_2 are the number of leaves belonging to nodes i, j

The output of the presented clustering algorithm is a binary tree (Fig. 4), this tree is transformed to an application interaction hierarchy based on the degree of similarity between the applications (Fig. 5).

4 Navigating between applications based on the user's application of interest

Based on the application interaction hierarchy produced in section 3 and the textual content provided by the user, the system interacts with the user to determine the target application. This process is described in detail in the following algorithm:

1. **Start**
The system starts with a generic prompt: "What can I do for you?" (similar to the internal GDN "Start" described in [1]).
2. **Active node determination**
When receiving a user request (represented by a vector q), the system determines the active node on the hierarchy based on the following steps [5]:
 - (a) **Score computation**
Compute the similarity between q and d_0, d_1, \dots, d_{n-1} , we obtain a set of scores s_0, s_1, \dots, s_{n-1} : $s_i = sim(d_i, q)$.
For example, in Fig. 3 we have $s_0 = 0.85, s_1 = 0.9, \dots, s_9 = 0.15$.
 - (b) **Upward propagation**
Select the best scores at each level and propagates them upward until the root is reached.
For example, in Fig. 3 we have $s_{0-2} = \max(s_0, s_1, s_2) = 0.9$.

- (c) Downward traversal to determine the active node
 Start from the root, compute the difference between two highest score child nodes, if this difference is below a certain threshold (we call this threshold the internal node stop threshold $t_s : (0 < t_s \leq 1)$), then stop. If not, go down to the highest score child node and continue to determine the active node.

For example, in Fig. 3, starting from M_{0-9} , we calculate the difference between M_{0-4} and M_{5-9} : $dif(M_{0-4}, M_{5-9}) = 0.4$, it is greater than $t_s = 0.15$, then we go down to M_{0-4} , we still have $dif(M_{0-2}, M_{3-4}) = 0.2$ is greater than t_s then we go down to M_{0-2} , we have $dif(M_0, M_1) = 0.05 < t_s$ then M_{0-2} is the active node.

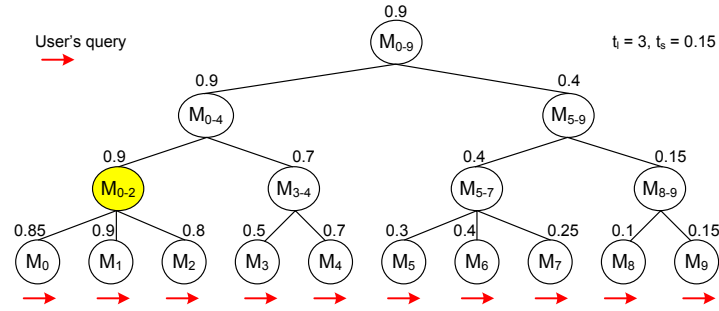


Fig. 3. Determine the active node based on the user's query

3. Response generation

The active node identified in the previous steps can be a root, an internal node or a leaf. Two types of response depending on the position of the active node:

- (a) The active node is the root or an internal node
 In this case, the functionality of the active node is similar to the list processing GDN described in [1]. The system shows a list of application candidates belonging to the active node and their score is not below the highest score leaf outside the active node (the maximum number of candidates shown to the user is limited by a threshold called the list processing threshold $t_l : (t_l \in N) \text{ and } (t_l \geq 1)$). The user can determine to go next (i.e. show the t_l following application candidates), previous (i.e. show the t_l previous application candidates), stop (i.e. restart the dialogue), up (i.e. move to the upper level on the hierarchy), down (i.e. move to the highest score child node), or select the desired application. In case, the user does not change the active node (i.e. he does not use the command up or down) and after browsing all the applications (belonging to the active node), he could not find his desired applications, the system

temporarily assigns the scores of the browsed leaves to zero and goes back to step 2b.

- (b) The active node is a leaf
 The application takes control and interacts with the user as an application-specific dialogue system. If the user request is out of the application's domain, go back to step 2.

An example of the algorithm with $n = 10$ and $t_l = 3$ is presented in Fig. 6 and explained in detail in section 5.3.

5 Scenario example

This section illustrates the process of developing a dialogue system accessing 10 applications in the ICIS domain using three steps presented in sections 2, 3, 4. The applications are: car route navigation (A_0), air route navigation (A_1), traffic lanes (A_2), map and fire management (A_3), tunnel sensors management (A_4), weather forecast (A_5), virtual control room (A_6), road surface temperature monitoring (A_7), patient information search (A_8), and medical worker verification (A_9).

5.1 Step 1

Applying the RDPM, we produce the finalized dialogue models: M_0, M_1, \dots, M_9 .

5.2 Step 2

From the finalized dialogue models, we create the application interaction hierarchy (cf. Fig. 5). Finalized dialogue models for the root and internal nodes are the list processing mGDNs produced by the RDPM. The role of each node is to select a subset of the applications belonging to it, for example the role of M_{0-2} is to select a subset of $\{A_0, A_1, A_2\}$.

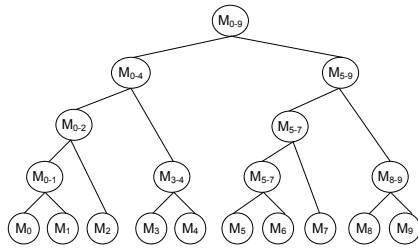


Fig. 4. Binary tree

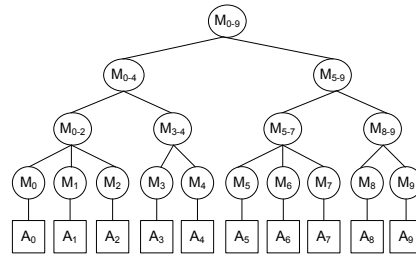


Fig. 5. Application Interaction Hierarchy

5.3 Step 3

An example of the system-user interaction is presented in Fig. 6. The “Start” mGDN sends the system’s prompt S_1 to the user. According to the content of the user’s prompt U_2 , the active node M_{0-2} is determined. M_{0-2} asks the user to select an application from the list $\{A_0, A_1, A_2\}$. Based on the user’s answer in U_3 , M_0 is activated. In steps from 5 to $k-1$ M_0 interacts with the user as an application-specific dialogue system. In step k , the user’s request U_k is out of M_0 ’s application domain, M_0 then forwards U_k to the system. The system analyzes U_k and activates M_8 . M_8 continues the interaction with the user in a similar manner M_0 has done.

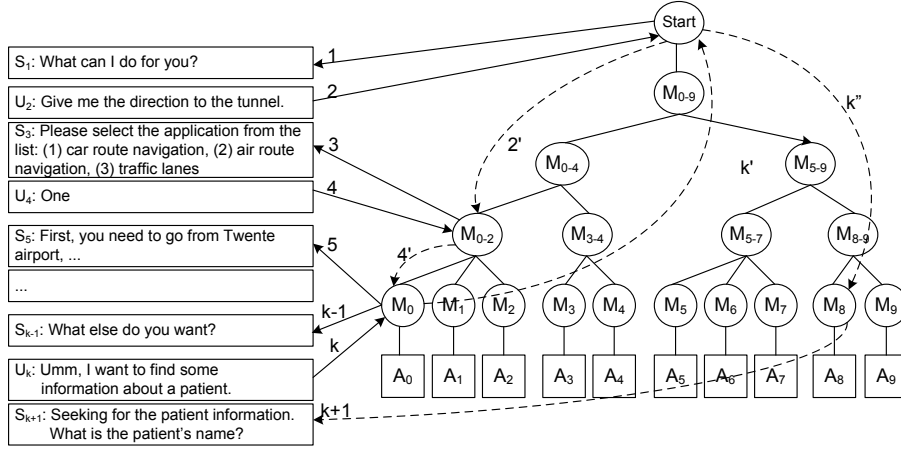


Fig. 6. Navigating between the applications

6 Conclusion

We have presented a framework for the development of interfaces for multi-application dialogue systems. Three important steps in the framework have been described and illustrated by a scenario example.

Currently, the RDPM software toolkit is available for the development of finalized dialogue models for single applications. It has been used in three re-

search projects: InfoVox ⁴, INSPIRE ⁵, IM2.MDM ⁶ and is being extended for the development of a large number of applications in the ICIS domain. The practical result shows that from a simple application, we can develop an initial dialogue model in some hours. The dialogue manager, the most important part of dialogue prototyping, covers most of the application independent dialogue functionalities (i.e. branching logic, dialogue dead-end management strategy, confirmation strategy, dialogue termination strategy, incoherencies, strategy defining level of initiative, etc.) Therefore, we can re-use the dialogue manager and the other modules described in section 2.3 for the development of multi-application dialogue systems.

Some initial work toward developing the application interaction hierarchy and navigating between the applications (sections 3, 4) has been analyzed and implemented (e.g. NLP Pre-Processing Tool, VS model). The multi-application dialogue system for ICIS domain presented in section 5 is currently under development.

7 Future work

Two main possible extensions of the generic dialogue modeling methodology we plan to study in the future: crossing-application and task selection.

7.1 Crossing-application

The application interaction hierarchy created in section 3 can be used to manage several concurrent applications (i.e. crossing-application). This is the case when the system goes to an active node containing a small number of applications. This extension is significant when the user wants to simultaneously execute several applications in order to achieve his goal in an optimal way.

7.2 Task selection

In the definition of the application interaction hierarchy in the section 3.1, we mentioned that each leaf corresponds to an application. In task-oriented dialogues, each application usually consists of several tasks. We can extend the hierarchy for identifying a task or a set of tasks in an application. To achieve this goal, the hierarchy will be constructed from the set of tasks in the same way we have done for the set of applications. Further work on task sharing (i.e. one task appears in several applications) will be studied.

⁴ The purpose of the InfoVox project was to prototype a dialogue-based vocal information server for the access, over the phone, to information about the restaurants of the city of Martigny (Switzerland). The project was partly funded by the Swiss national CTI grant program.

⁵ The INSPIRE project aiming at a dialogue based control of various home devices (lights, TV, VCR, ...) within a Smart Home environment, was funded by the European FP5 IST research grant program.

⁶ <http://www.issco.unige.ch/projects/im2/mdm/>

8 Acknowledgements

We would like to thank Martin Rajman for his useful suggestions and comments about the generic dialogue modeling idea presented in this paper. This work is part of the ICIS program (<http://www.decis.nl/html/icis.html>). ICIS is sponsored by the Dutch government under contract BSIK 03024.

References

1. Trung H. Bui, Martin Rajman, and Miroslav Melichar. Rapid dialogue prototyping methodology. In *Proceedings of the 7th International Conference on Text, Speech & Dialogue (TSD 2004)*, P. Sojka, I. Kopecek & K. Pala (Eds.), *Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg New York pp. 579-586, Brno, Czech Republic, September 8-11 2004.
2. Trung H. Bui, Martin Rajman, and Silvia Quarteroni. Extending the Rapid Dialogue Prototyping Methodology: User Modeling. Technical Report WP4 Deliverable 4.3, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, August 2004.
3. Jennifer Chu-Carroll and Bob Carpenter. Vector-based natural language call routing. *Computational Linguistics*, 25(3):361-388, 1999.
4. Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. Scatter/gather: a cluster-based approach to browsing large document collections. In *SIGIR '92: Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 318-329, New York, NY, USA, 1992. ACM Press.
5. Edouard Forler. *Intelligent user interface for specialized web sites*. Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, diploma thesis edition, 2000.
6. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264-323, 1999.
7. S.C. Johnson. Hierarchical Clustering Schemes. *Psychometrika*, pages 2:241-254, 1967.
8. Agnes Lisowska, Martin Rajman, and Trung H. Bui. Archivus: A system for accessing the content of recorded multimodal meetings. In Samy Bengio and Hervé Bourlard, editors, *Machine Learning for Multimodal Interaction, First International Workshop, MLMI 2004, Martigny, Switzerland, June 21-23, 2004, Revised Selected Papers*, volume 3361, pages 291 - 304. Springer.
9. Hong-I Ng and Kim-Teng Lua. Dialog Input Ranking in a Multi-Domain Environment Using Transferable Belief Model. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan, July 2003.
10. M. Rajman, R. Besanon, and J.-C. Chappelier. Le modèle DSIR : une approche à base de sémantique distributionnelle pour la recherche documentaire. *revue Traitement Automatique des Langues (TAL)*, 41(2), Paris, 2001.
11. Bor shen Lin, Hsin min Wang, and Lin shan Lee. Consistent dialogue across concurrent topics based on an expert system model. In *Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, Budapest, Hungary, 1999.
12. Jürgen te Vrugt and Thomas Portele. Application-Independent Knowledge-Processing in a Task-Oriented Speech-Dialog-System. *it - Information Technology*, 46(6):306-314, December 2004. <http://www.it-inftech.de>.