GOSSIP GALORE
A Conversational Web Agent for Collecting and Sharing Pop Trivia

Feiyu Xu, Peter Adolphs, Hans Uszkoreit, Xiwen Cheng and Hong Li

DFKI GmbH, Language Technology Lab, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany

Keywords: Web mining, Relation extraction, Web intelligence, Intelligent user interface, Conversational agent, Question answering, Dialogue system.

Abstract: This paper presents a novel approach to a self-learning agent who collects and learns new knowledge from the web and exchanges her knowledge via dialogues with the users. The application domain is gossip about celebrities in the music world. The agent can inform herself and update the acquired knowledge by observing the web. Fans of musicians can ask for gossip information about stars, bands or people and groups related to them. This agent is built on top of information extraction, web mining, question answering and dialogue system technologies. The minimally supervised machine learning method for relation extraction gives the agent the capability to learn and update knowledge constantly from the web. The extracted relations are structured and linked with each other. Data mining is applied to the learned data to induce the social network among the artists and related people. The knowledge-intensive question answering technology enhanced by domain-specific inference and active memory allows the agent to have vivid and interactive conversations with users by utilizing natural language processing. Users can freely formulate their questions within the gossip data domain and access the answers in different ways: textual response, graph-based visualization of the related concepts and speech output.

1 INTRODUCTION

The development of information extraction and question answering in recent years opens new perspectives for simple but effective interactive dialogue systems (Jönsson et al., 2004; Strzalkowski et al., 2005; Theune et al., 2007). Information extraction enables dialogue systems to access and understand natural language texts stored in semi- or unstructured formats, thus, allowing them to make use of the contents provided by the web, the world's largest information repository. Question answering technology gives a conversational agent the capability of understanding natural language questions and retrieving answers from a large knowledge or content pool. At the same time, question answering systems enhanced by some dialogue competence enable natural communication with the human users. The combination of information extraction, question answering and dialogue is a new approach to a conversational agent who is able to understand natural language questions and provide answers by extracting and mining information from a large amount of textual data in structured, semi- or

unstructured form.

One of the hardest challenges in our information world is to constantly keep the information up to date and to prepare it in such a way that users can easily understand and exploit it. We have developed a new architecture for conversational agent systems that can learn, update and interpret information from the web and make conversations with end users, provide answers to their questions and even help them to gain insights into the application domain. We selected gossip about celebrities in the music world as the domain for our experimental setup, because many of them exhibit interesting and dynamic aspects with respect to both their private and professional life. Furthermore, they are connected to each other in a variety of ways. Internet news and blogs report on them from different perspectives. Our task is to model this domain by covering relevant facts and trivia on the musicians and their communities and by discovering new properties and relations. The acquired information will be utilized as a knowledge resource for conversations with end users. Users can raise natural language questions about a special artist or ask for relationships between

artists. Our system provides answers from its knowledge base or even hints at newly discovered information.

In comparison to existing systems, our conversational agent, called "Gossip Galore", is an active self-learning system. It starts with only a very small number of artists and bands and then gradually finds many more artists and bands. This is realized by the application of a minimally supervised relation extraction system (see section 3). Users can actively give comments on the answers provided by the agent, which is useful for self validation. Thus, "Gossip Galore" contains two major parts: one is the knowledge acquisition component and the other one is the component for communication and conversation. Both parts interact with each other and contribute to the self-learning process.

The paper is structured as follows: Section 2 describes the project and the general context in which "Gossip Galore" is embedded. Section 3 explains the web mining techniques for the knowledge acquisition, while section 4 presents the dialogue modelling and question answering component. Section 5 gives an overview of the related work. The conclusions and future steps are described in section 6.

2 RASCALLI

The research presented here is conducted within the project *Responsive Artificial Situated Cognitive Agents Living and Learning on the Internet* (RASCALLI). RASCALLI is supported by the Sixth Framework Programme of the European Commission in the area of Cognitive Systems (IST-27596-2004). Its goal is to develop and implement cognitively enhanced artificial companions by combining natural language processing, question answering, web-based information extraction, semantic web technology and interaction-driven profiling with cognitive modelling (Krenn, 2008). This work is further supported by the project KomParse, which is devoted to equipping non-player characters in computer games with dialogue capacities.

In the realized system, the RASCALLI agents assist users in extracting information from the web and other resources. Users can own their own RASCALLI agents, which are 3D modelled virtually embodied conversational agents. The perception and action components of the RASCALLI agents are modelled by a combination of information extraction, question answering and dialogue capabilities. Within the project some major strands of research are devoted to the investigation and modelling of architectures that

combine all major components of cognitive systems. This is an ambitious and demanding task, and as a step on the way, the results reported here are a pragmatic compromise that combines state-of-the-art and novel methods from information extraction, question answering, semantic technologies and visual animation with insights from cognitive modelling into a robust fun application.

3 WEB MINING FOR KNOWLEDGE ACQUISITION

One of the major competences of the RASCALLI agents is that they can learn and acquire knowledge constantly from the web according to user interests. The minimally supervised machine learning methods for relation extraction provided by the system DARE can be easily utilized for realizing this competence (Xu et al., 2007; Xu et al., 2008a). DARE can be initialized with several examples of relations about artists or bands as seed provided by the users and then learn rules which map the linguistic structures to these semantic relations. The rules can be applied to texts to discover new relation instances, which can be reused as seed again for new rule discovery.

The experimental domain selected for RASCALLI is gossip about celebrities in the pop world. We start with domain modelling to define the potentially relevant concepts and relations that will serve as a framework for the musician profiles and the associated gossip information to be acquired. Given the relevant concepts and their relations, we apply DARE to acquire instances of the relations from the web.

3.1 Domain Modelling

The aim of the domain modelling is to identify and structure the relevant concepts and relations within the gossip domain. The current domain contains properties of a musician such as personal profiles, social contexts, achievements, gossip topics and career relevant issues. The gossip content is modeled as an ontology, utilizing the formal language OWL. The concepts and properties centering on the musician are depicted in figure 1.

3.2 Knowledge Acquisition

Many resources on the web report on celebrities, e.g., online news sites, Wikipedia, music portals, fan blogs and forums. The information mentioned above is stored in different formats: unstructured (free text), semi-structured (e.g., Wikipedia) or almost structured
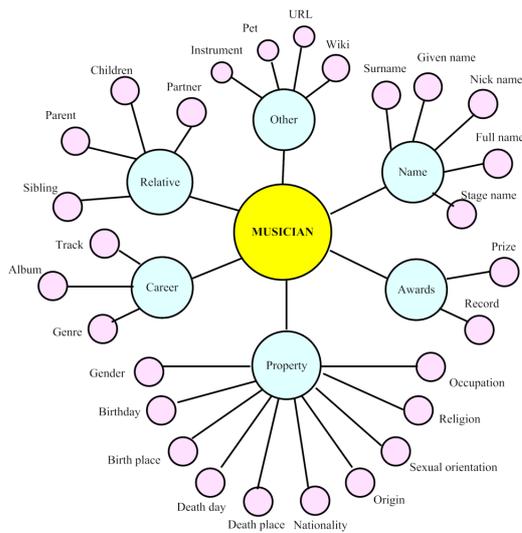
Figure 1: Domain ontology (simplified).
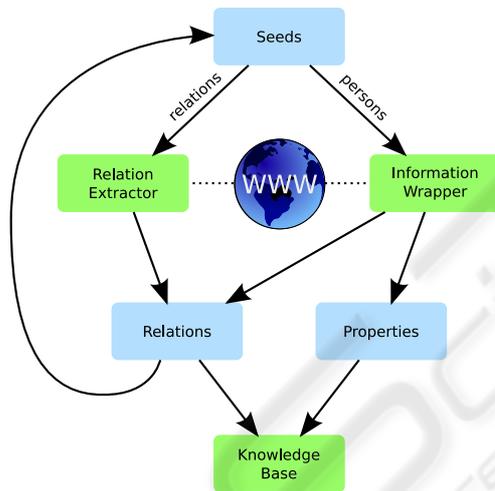
Figure 2: Webmining workflow.

(e.g. NNDB). Therefore, we propose a hybrid information discovery strategy to detect as much information as possible, as shown in figure 2.

We apply information wrapping, information extraction and information merging techniques to acquire new knowledge. The whole discovery is embedded in a bootstrapping framework, namely, starting with some examples and then learning more and more information after several iterations.

Relation Extraction with DARE. DARE provides a general framework for the extraction of relations and events with various complexities (Xu et al., 2007; Xu et al., 2008a). This method is minimally supervised since the system works with a collection of free

natural language texts without any annotation of domain information. The only domain knowledge for the whole process is the seed. DARE can use linguistic knowledge as it is provided, for example, by named entity recognizers and linguistic parsers. The complexity of the seed determines the complexity of the extracted relations. The seed helps us to identify the explicit linguistic expressions containing mentions of n-ary relation instances or instances of their k-ary projections where $1 \leq k < n$. Therefore, DARE can be easily adapted to user interests. Users provide only some new examples of the relations they are interested in; DARE can learn additional information from the web based on these examples.

In the current system, we apply SProUT (Drozdzynski et al., 2004) for the recognition of person names and other concepts (e.g., band and group names, date time, nationalities, instruments, religions, sexual orientations) and utilize the Stanford Parser (Klein and Manning, 2003) to detect linguistic dependency structures. DARE was originally used to extract information about Nobel Prize winners from free text. Later experiments showed how to adapt learned DARE rules for the Nobel Prize award domain to discover awards won by musicians (Xu et al., 2008b). Let us look at the following example. Given a seed example about a *Grammy* award won by *Madonna* for a specific category in the year *1992*:

(1) ⟨Madonna, Grammy, Best Long Form Music Video, 1992⟩

The natural language sentence which matches this seed is:

(2) *Madonna* won her first *Grammy* in *1992* in the *Best Long Form Music Video category* for the laserdisc release of her 1990 Blond Ambition Tour.

DARE can extract a linguistic pattern from the seed example and the matched sentence where the linguistic arguments are associated with their semantic roles in the semantic relation, after applying linguistic analysis to the sentence. The simplified DARE rule looks as follows:

(3) ⟨subject: recipient⟩ <u>win</u> ⟨object: prize⟩ ⟨mod: year⟩ ⟨mod: category⟩

Information Wrapping. Information wrapping is responsible for collecting structured data from structured or semi-structured web sites. It discovers the HTML structures which indicate the relations defined in our ontology. We apply this technology to web sites such as Wikipedia and the special web portal for people and their profiles, namely, the NNDB.
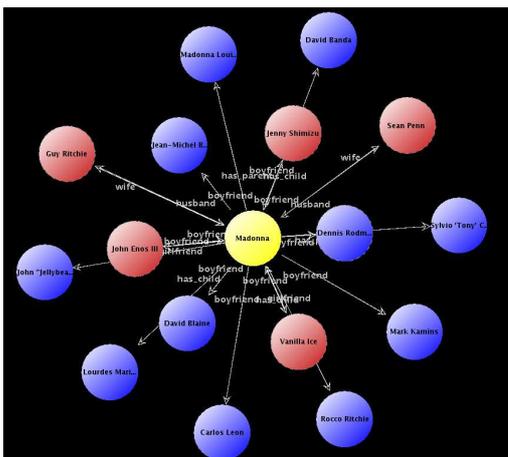
Figure 3: Social network of Madonna.

The method starts with a set of musicians and their relation instances as seed. Our system sends a query containing an instance from the seed set as a query to the web sites and discovers the rules which map the HTML structures to the relation structures.

Induction of the Social Network.   Given the discovered relations among the musicians themselves and other people, we developed a special system which can construct a social network from the relation instances. For example, figure 3 shows the social network of Madonna. The social network also serves as the basis for the active dialogue memory of the agent.

4 CONVERSATIONAL AGENTS

In RASCALLI, the central method for users to access the acquired knowledge is to communicate with the user's personal embodied conversational agent (ECA). The core functionality of the agent is question answering, wrapped in a smooth natural language dialogue. One main design criterion is to create and enhance an immersive effect on the user when interacting with the system. The agent should be physically embodied, she should be situated in a consistent physical environment, and she should act naturally.

The interaction between users and RASCALLI can be described as follows. After logging in to the platform, a three-dimensional visualization of the user's agent is displayed (see figure 4). Just as in an instant messaging program, the user communicates with the agent by typing messages into a text field. The agent, on the other hand, responds with natural language utterances which are presented in their spoken form (by the use of the open source speech
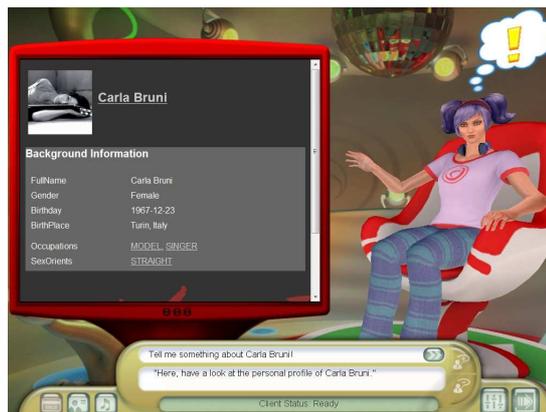
Figure 4: The Gossip Galore conversational agent.

synthesis system *OpenMary* (Schröder and Hunecke, 2007)), along with their written form. But the agent's means of communication are not restricted to verbal actions. Complex answers such as the social network of a star can be visualized on a TV screen, which is embedded into the agent's environment. Where it is appropriate, the agent also emphasizes her responses by facial and body gestures such as shrugging the shoulders, nodding or shaking her head and pointing to the screen.

4.1 Architecture

The RASCALLI system is realized as a server-client architecture. Users are connected to the server via the 3D client, which displays our ECA and manages the interaction with the user. The actual control logic of the agent is executed on the server. The server's function is to accept new connections, to manage users and their logins, and to route messages between the 3D client and the conversational agent.

Figure 5 shows the component hierarchy of our conversational agents. Some of the components are responsible for processing various linguistic aspects of the dialogue, whereas others are concerned with knowledge representation, management and retrieval as well as behavioural procedures. Details on the interplay between these components when processing dialogue turns are presented in the following subsection.

4.2 Dialogue Processing

When the conversational agent receives an utterance from the user, its task is to compute a suitable dialogue turn in response. We follow a pipeline architecture for realizing this: the user's input string is first linguistically analyzed, then it is interpreted in
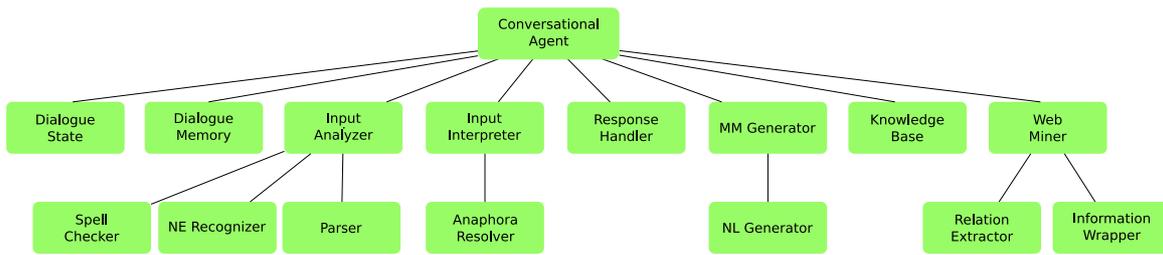
Figure 5: Components of a conversational agent.
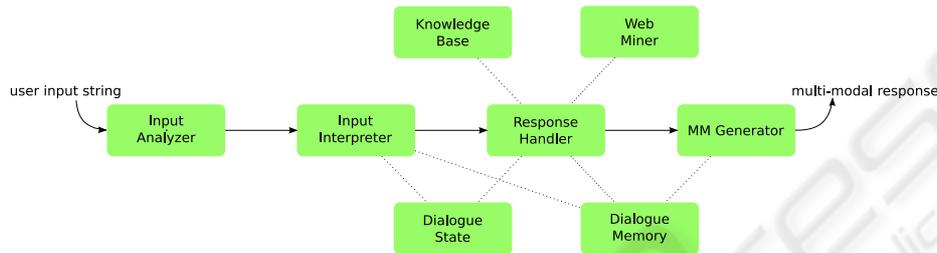
Figure 6: Processing dialogue turns.

the current dialogue context and turned into a suitable plan for a response action that is executed in the third stage, leading to an abstract representation for the answer, which is realized with verbal and non-verbal means in the fourth and final stage. This basic data flow when processing dialogue turns as a response is depicted in figure 6. In the following, the four main components are presented in greater detail.

The idea of having two separate components for input processing, namely, *input analyzer* and *input interpreter*, one for the analysis and one for the interpretation of the user's input, serves the purpose of drawing a clear boundary between the general and reusable and the domain-specific parts of the system. The input analysis component relies on standard domain-independent linguistic tools, namely a spell checker, a named entity recognizer, and a parser producing a linguistic analysis of the input, for which we currently employ a fuzzy paraphrase matcher to approximate the output of a deeper syntactic/semantic parser.

Each utterance is associated with a meaning representation as well as with the answer focus and the expected answer type in case of questions. Note that by mapping utterances of quite different sentence types such as plain questions ("Who is Madonna?"), statements with embedded questions ("I wonder who Madonna is."), statements about the user's interests without embedded questions ("I'm interested in Madonna.") to the same semantic representation, we can conflate sets of user utterances with the same intended meaning.

In the second stage, the input is interpreted in the current dialogue context, considering previously men-

tioned entities for resolving anaphora as well as the current dialogue state for modelling the system's expectations about the user's turn. If, for instance, a substring can only be resolved as a named entity with the help of the spell checker, the system poses a clarification question to the user and sets the dialogue state accordingly. This allows the system to interpret a following utterance by the user such as "yes" or "no", which would otherwise not be understood. The result of the input interpretation stage is an abstract plan to perform a certain action. For example, factoid in-domain questions result in a plan to look up the data in our knowledge base, general information requests about an in-domain person result in the plan to show a profile page of that person, out-of-domain questions about a known person result in the plan to present a suitable web link, and so on.

In the third stage, the Response Handler component executes the planned action. For factoid questions, this means that the corresponding query is looked up and submitted to the knowledge base. If the user asks for general information about a person, the URL of the corresponding profile page is constructed. The user may also have asked whether there are any new information about a musician he is interested in. In this case, an online search for new connections between people is performed.

The agent follows certain pragmatic principles of relevance when giving answers to questions. By assigning the same semantics to indirect speech acts ("I wonder who the boyfriends of Madonna are.") as to the corresponding direct speech act ("Who are the boyfriends of Madonna?"), we are able to return a rel-

evant response to the user's request. Similarly to the previous example, certain yes/no-questions ("Does Madonna have any boyfriends?") can be answered as if they were wh-questions. Instead of giving a simple "yes" answer, the agent also lists the values for the queried variable.

Not all of the performed actions necessarily lead to a satisfying result, though. If no positive answer can be found for a question or if the question lies outside the covered domain, we still want to be able to provide a constructive answer. If, for instance, the user asks about a person we do not have detailed information about but for whom a Wikipedia entry exists (e.g. "Tell me something about Nicolas Sarkozy!"), we point the user to this page, using the embedded TV screen for displaying the page. If, on the other hand, the user asks an in-domain question for which the system does not have any results, we direct the user to a Google search page with appropriate query parameters in order to help him find relevant information.

The outcome of the performed action is always an abstract representation of the agent's response. This might be a simple boolean value for yes/no-questions, a list of entities for factoid questions, a URL for the system's own information services or to external sources, and so on. This information is finally realized as a communicative act in the fourth stage, the multi-modal generation. We currently employ template-based generators for both producing the natural language utterance as well as for the multi-modal message with gestures and TV screen commands. When generating natural language answers to questions, care has to be taken how these answers are provided. Since we have to expect that our knowledge base is incomplete and that the acquired information could partly be inaccurate (particularly in the gossip domain), special relativizing expressions such as "according to my sources" are produced as part of the answer. The introduction of pronouns for entities mentioned before helps making the utterance less static and the conversation more natural.

4.3 Multimodal Communication

Gossip Galore uses several modalities for communicating with the user. First of all, all the agent's utterances are spoken, with the help of a speech synthesis system. The verbal part of the answer is additionally supported by gestures. To make maximal use of the available means for communication, we also use a TV screen embedded in the agent's environment, which is able to display arbitrary web sites, to present illustrations of the current answer (see figure

Figure 7: Multimodal answers – illustrating locations.

Figure 8: Social network visualization.

7) or even to provide the very content of the answer where the answer is not a single fact or a small set of facts but would require a complex explanation involving heterogenous kinds of information (see figure 4) or would lead to a rather longish and tiring answer if it were realized verbally (see figure 8).

5 RELATED WORK

Web-based question answering systems typically proceed in several stages: i) the question is turned into a query for a standard search engine, ii) a set of relevant web sites is retrieved, iii) text passages are selected as an answer from that document set (Neumann, 2008). A variation of this idea is applied in HITIQA (Strzalkowski et al., 2005), an interactive open-domain question answering system for complex exploratory questions, where the answer is retrieved based on complete semantic event frames which are matched against the frame of the question in the last of

the three steps above. Thus, much like in RASCALLI, the system performs a more structured semantic analysis of the original data (with respect to the question at hand). In contrast to our system, and more in accordance with the common web-based QA approach, however, the extracted information is only used for selecting the most relevant text passages. Furthermore, it is only used for the current QA task, not for building up world knowledge.

The approach followed in our system is to learn new information from a large unstructured text pool and store it in a knowledge base for structured access. Such an approach is also followed in BIRDQUEST (Jönsson et al., 2004), a QA system for answering ornithological questions. As in our system, the information is extracted from natural language text (although from a single source – a bird encyclopedia – with much stricter conventions than arbitrary web documents) and then stored in a relational database. Unlike our system, however, the system is not self-learning; suitable information extraction patterns are not learned automatically but have to be provided as a resource.

The potential benefits and sub-tasks involved in enhancing question answering with dialogue capacities to get interactive question answering have been briefly discussed in the Q&A Roadmap (Burger et al., 2000). Recourse to a discourse memory for tracking entities over several questions has played a role in the context task of the QA track at the TREC 2001 conference and when processing question series in the main tasks in the QA tracks of TREC 2004 and 2005. The extension of question answering to more interactive dialogue has been tackled in the Complex Interactive QA (ciQA) task in the QA tracks of TREC 2006 and 2007[1].

There are several projects that enhance a QA system with more interactive capabilities, namely BIRDQUEST (Jönsson et al., 2004), HITIQA (Strzalkowski et al., 2005), RITEL (Rosset et al., 2006), the IMIX demonstrator (Theune et al., 2007), and SMARTWEB (Reithinger et al., 2007). RASCALLI has a different focus compared to all of these systems in that it i) aims to create a personal relationship with the user by the use of user-adaptive knowledge acquisition methods, and ii) conducts a vivid conversation with the user that mimics human-to-human communication, creating the immersive effect of a living entity with its own personality.

---

[1]Please refer to the TREC homepage at http://trec.nist.gov/ for further information and references.

6 CONCLUSIONS

We have described the overall architecture and main components of a new class of web-based virtual agents. Although the design of the agents is strongly influenced by empirical observations and theoretical models of natural cognitive agents, our goal has not been a simulation of biological cognition. This aim is partially targeted by other research strands within the RASCALLI consortium. The objective of the demonstrated architecture and implementation has been a rather pragmatic and simplified agent model that exhibits the desired performance properties and serves as the starting point for a range of extensions and additional applications. The achieved relevant performance properties are: robustness, accuracy, self-improvement and nearly real-time behavior.

The planned future extensions include the integration of deeper language processing methods instead of or in addition to the fuzzy paraphrase matcher. A prime candidate for this extension is our own deep syntactic/semantic parser. Another plan concerns the required temporal aspects of relations. It is the dynamics of the domain that provide the basis for the gossip. Properties and relationships change quite often. By detecting and relating the utterance and report times of the various information sources, a multitude of answers may be temporally sorted. Once in a while, contradicting information is harvested. In some cases, these contradictions result from an unresolved temporal succession, i.e. the contradicting facts were true at different times. In other cases, one of the contradicting facts is simply false. In order to deal with such situations, we need to enrich the information extraction by methods for credibility checking, which will be adopted from IE/IR research.

Finally, we plan to exploit the dialogue memory for moving more of the dialogue initiative to the agent. In cases of missing or negative answers or in cases of pauses on the user side, the agent can use the active parts of the dialogue memory to propose additional relevant information or to guide the user to fruitful requests within the range of user's interests. However, the hardest test for the agent architecture will be the extension to other domains and tasks that may be less error forgiving than the colorful world of pop trivia.

REFERENCES

Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., Jacquemin, C., Lin, C.-Y., Maiorano, S., Miller, G., Moldovan, D., Ogden, B., Prager, J., Riloff, E., Singhal, A., Shrihari, R., Strzalkowski, T., Voorhees, E., and Weishedel, R. (2000). Issues, tasks and program structures to roadmap research in Question & Answering (Q&A).

Drozdzynski, W., Krieger, H.-U., Piskorski, J., Schäfer, U., and Xu, F. (2004). Shallow processing with unification and typed feature structures – foundations and applications. *Künstliche Intelligenz*, 1:17–23.

Jönsson, A., Andén, F., Degerstedt, L., Flycht-Eriksson, A., Merkel, M., and Norberg, S. (2004). Experiences from combining dialogue system development with information extraction techniques. In Maybury, M. T., editor, *New Directions in Question Answering*, pages 153–168. MIT Press.

Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 423–43.

Krenn (2008). Responsive artificial situated cognitive agents living and learning on the internet. Poster presented at the International Conference on Cognitive Systems (CogSys 2008).

Neumann, G. (2008). Strategien zur Webbasierten Multilingualen Fragebeantwortung: Wie Suchmaschinen zu Antwortmaschinen werden. *Computer Science - Research and Development*, 22(2):71–84.

Reithinger, N., Herzog, G., and Blocher, A. (2007). SmartWeb – mobile broadband access to the semantic web. *KI – Künstliche Intelligenz*, 2/2007.

Rosset, S., Galibert, O., Illouz, G., and Aurélien, M. (2006). Integrating spoken dialog and question answering: the Ritel project. In *Proceedings of INTERSPEECH 2006*.

Schröder, M. and Hunecke, A. (2007). Mary tts participation in the Blizzard Challenge 2007. In *Proceedings of the Blizzard Challenge 2007*, Bonn, Germany.

Strzalkowski, T., Small, S., Hardy, H., Yamrom, B., Liu, T., Kantor, P., Ng, K., and Wacholder, N. (2005). HI-TIQA: A question answering analytical tool. In *Proceedings of the International Conference On Intelligence Analysis (IA-05)*, McLean, VA.

Theune, M., Krahmer, E., van Schooten, B., op den Akker, R., van Hooijdonk, C., Marsi, E., Bosma, W., Hofs, D., and Nijholt, A. (2007). Questions, pictures, answers: Introducing pictures in question-answering systems. In *Tenth international symposium on social communication*, pages 450–463, Cuba.

Xu, F., Uszkoreit, H., and Li, H. (2007). A seed-driven bottom-up machine learning framework for extracting relations of various complexity. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 584–591.

Xu, F., Uszkoreit, H., and Li, H. (2008a). Task driven coreference resolution for relation extraction. In *Proceedings of the European Conference for Artificial Intelligence ECAI 2008*, Patras, Greece.

Xu, F., Uszkoreit, H., Li, H., and Felger, N. (2008b). Adaptation of relation extraction rules to new domains. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.