9-2014

Learning Situated Knowledge Bases through Dialog

Aasish K. Pappu
Carnegie Mellon University

Alexander I. Rudnicky
Carnegie Mellon University, ar28@andrew.cmu.edu

swer a query. This knowledge might include the user's previous search behavior, emails and other information about the user [1]. Such knowledge is not readily available in a general commonsense knowledge base or a domain knowledge base for a typical dialog agent. To this end, we aim to define dialog strategies that an agent can use to build up a *knowledge base* based on information obtained through interaction with users.

Collaborative effort from people can help an agent solve complex problems. For example [10] have shown that people can build a commonsense knowledge base by playing games with a computer. Our work aims at building a knowledge base through purposeful dialog between a system and its users. [11] have shown that a variety of information can be obtained: for example, people can provide answers to visual questions and aid physically-disabled users. In our work, we seek to acquire knowledge available only through users to help the system provide better quality responses to subsequent user queries. We address the following questions:

1. *Can dialog-driven acquisition capture domain knowledge?* The agent solicits information from its users according to dialog strategies. This information is used to augment a knowledge base for the domain. We evaluate the coverage provided by this knowledge base both qualitatively and quantitatively.

2. *Is the acquired knowledge useful to the system?* The knowledge acquisition process should aid the system to improve its task success rate. To this end, we evaluate the system's performance in its event information-access domain, before and after knowledge acquisition. We show that acquired knowledge significantly improves the system's performance, as assessed by independent judges.

This paper is organized as follows. First, we describe knowledge acquisition strategies and a user study where system acquires information from users using these strategies. In Section 3, we describe how agent learns a knowledge base from the acquired information. We also present a qualitative and quantitative analysis of the knowledge base. In Section 4, we evaluate the performance of the system on an information-access task, to show that acquired knowledge is indeed useful for system performance. Finally, we make concluding remarks and brief about future directions to this work.

2. Situated Knowledge Acquisition Study

We posit three different ways in which a dialog agent can interactively acquire new information from users:

(I) Knowledge acquired through explicit user guidance. For example, a human points at a blue box and tells the agent *this*

Abstract

To respond to a user's query, dialog agents can use a knowledge base that is either domain specific, commonsense (e.g., NELL, Freebase) or a combination of both. The drawback is that domain-specific knowledge bases will likely be limited and static; commonsense ones are dynamic but contain general information found on the web and will be sparse with respect to a domain. We address this issue through a system that solicits situational information from its users in a domain that provides information on events (seminar talks) to augment its knowledge base (covering an academic field). We find that this knowledge is consistent and useful and that it provides reliable information to users. We show that, in comparison to a base system, users find that retrievals are more relevant when the system uses its informally acquired knowledge to augment their queries.

Index Terms: Spoken Dialog, Knowledge Base, Situated Domain, Human Knowledge

1. Introduction

In many information access applications, systems need to update their domain knowledge over time to maintain accuracy. For example, an event-recommendation agent would need to know when new events appear in its domain. General information is available from the web. For example commonsense knowledge bases such as NELL [1], Freebase [2]. The agent can find information in these knowledge bases, but only about popular entities. Other types of information are also available, for example about popular concerts, but only those with a web presence. We are interested in acquisition of the latter type of knowledge, not always present on-line but shared on an informal basis within groups, and how it can be obtained through interaction with people.

Knowledge bases such as NELL, Freebase, and Wordnet can help in expanding semantic context, thus improve text classification in low training data scenarios [3, 4, 5]. In our previous work [6], we found that these knowledge bases are useful in improving dialog task prediction by expanding a user query with additional semantic context. However, the semantic context is only applicable to common content words as opposed to specific entities in the domain e.g., names of specific events. On the other hand, people that might interact with a dialog system providing information access in a domain, can provide knowledge which is useful. The dialog agent could therefore proactively seek information from its users and build a *knowledge base* or a *folksnomy* for its domain. This process is sometimes known as collaborative tagging [7]. As a result, the agent gains access to ontologies of information that are present in the users' minds.

Building knowledge bases is essential for a system that needs to answer user queries with situated and contextual information. Systems such as Google Now [8], Apple Siri [9] use a variety of secondary knowledge sources about the user to an-

---

Table 1: System initiated strategies used by the agent to drive knowledge acquisition

StrategyType	Strategy	Example of System's Prompt
Query Driven	*QueryEvent*	I know events on campus. What do you want to know?
	QueryPerson	I know some of the researchers on campus.Whom do you want to know about?
Egocentric	*Buzzwords*	What are some of the popular phrases in your research?
	FamousPeople	Tell me some well-known people in your research area
Show & Ask	*Tweet*	How would you describe this talk in a sentence, say a tweet.
	Keywords	Give keywords for this talk in your words.
	People	Do you know anyone who might be interested in this talk?

is a blue box. This form of learning assumes that human is an expert user of the system and knows exactly what the agent does not know. In a domestic robot domain [12, 13] have shown that users can teach novel locations to robots using a combination of trigger phrases and non-verbal gestures directed at the robot. [14] have demonstrated that system can learn one-word descriptions for objects in the environment. It has been shown that expert users can teach route plans to a mobile robot [15, 16].

(II) Knowledge acquired through "non-understanding" exchanges; i.e., the system cannot extract expected slot-values from an utterance [17]. For example, the system detects that there is an out-of-vocabulary(OOV) entity in the utterance, then tries to clarify with user and learn this entity. [18, 19] have proposed speak-spell strategy to acquire spellings of OOV words through interaction. [20, 21] has shown that a system can detect and learn new words in a speech translation application.

(III) Knowledge acquired through elicitation. System initiates a dialog with user and elicits new information. In their Nanoklaus system, [22] used an exploratory strategy to ask questions about a concept introduced by the user. In [23], users are provided with a *saybox* to ask a factoid question and the system uses this question to create an agenda of questions probing unknown information related to this question. [24] has proposed *show&ask* strategies to elicit new information from people.

Based on the strategies proposed in [23, 24], we propose spoken dialog strategies to elicit information from users.

2.1. User Study Design

To investigate how people respond to system initiated knowledge acquisition, we conducted a user study. The study took place in the context of EventSpeak Dialog System that informs people about upcoming talks/events of their interest and ongoing work of other researchers on a university campus. This system takes a spoken query from user and responds with a set of talks related to the query. System has a database of 160 academic talks that are scheduled between April 2013 to May 2014 with metadata for each talk: title, speaker name, abstract, location and others.

To acquire situated knowledge, the agent uses the strategies shown in Table 1. In **Query Driven** strategies, system prompts user with an open-ended question akin to "How-may-I-help-you" to learn what "values" of a slot that user is interested in. This allows the system to respond to the query if it is familiar with the slot-value, otherwise it learns the unknown slot-value. In **Egocentric**, the system asks user about their own interests and people associated with those interests. This allows the system to learn information that people are already familiar with. In **Show & Ask**, system shows a description of an event and asks questions to ground user's responses in relation to that event. This allows the system to learn information in relation to an event. This strategy is similar to one strategy proposed in [24], where the user is expected to provide a title when

given a story description. Our study is web-based and involves recording of voice responses from people based on the system's prompt. We use Testvox[2] to display talk descriptions on web pages and Wami[3], a push-to-talk interface, to record user's spoken responses.

We recruited 40 graduate students from School of Computer Science at Carnegie Mellon, a representative sample of prospective users for EventSpeak system. Each subject performed spoken tasks involving *Query Driven*, *Egocentric* and *Show & Ask* strategies. System initiates a knowledge acquisition dialog with the user as illustrated in Table 1 and expects the user to respond with a spoken reply. The details of tasks are given below.

In the *Query Driven* tasks, system responds back to the user's spoken query in the *QueryEvent* strategy with a list of talks. System displays a list of talks if the slot-value in the query is familiar, otherwise it does not but learns the new information. In the *Egocentric* tasks, system uses *Buzzwords* strategy, asking the user about popular key-phrases in their research area. Then, system asks about well-known researchers (*FamousPeople*) in their area. In the *Show & Ask* tasks, system shows a description of the talk and asks three questions related to this description. It asks the user to give one sentence description about the talk (*Tweet*), key-phrases related to the talk (*Keywords*) and people who might be interested in the talk (*People*).

2.2. Corpus Collection

This user study yielded 64 minutes of audio data, with on average 1.6 minutes per participant. We have orthographically transcribed the user utterances. Then annotated the corpus[4] for people names, and research interests. Table 2 shows the number of unique slot-values found in the corpus. We observe that *Egocentric* task yielded relatively higher number of researcher names (the *FamousPeople* strategy) than other tasks. This may have happened due to people finding it easier to recall people names from their own research area, compared to names in other areas. Overall, the user study yielded 139 unique researcher names and 485 research interests.

Table 2: Breakdown of unique number of researcher names and researcher interests elicited/acquired by strategy type

StrategyType	Researcher Names	Research Interests
Query Driven	21	29
Egocentric	77	107
Show & Ask	76	390
Overall	139	485

Figure 1: Condensed version of the block partitioned network of researchers. Each vertex represents a block of researchers.
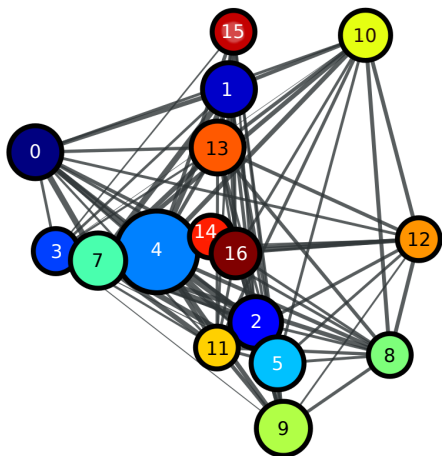

Figure 2: Research Interests associated with each block in the network. Blocks with similar interests have thicker edges.

```
0. speech synthesis, crowdsourcing
1. neural networks, graphical models
2. natural language processing, active learning
3. machine translation, social media
4. big data, active learning
5. information retrieval, distributed systems
6. high dimensional problems, sample complexity
7. speech recognition, human-computer interaction
8. clustering, applied machine learning
9. scalable optimization, structured sparse learning
10. protein structure, graphical models
11. information extraction
12. search, learning to rank
13. crowdsourcing, deep learning
14. community detection
15. deep learning, computer vision
16. information extraction, neuro science
```

3. Acquired Situated Knowledge Base

In this section, we address our first question: *Can the dialog-driven acquisition capture domain knowledge?* From the corpus collection, we have a list of researcher names and a list of research interests. To address our question, the system should infer a list of interests for each researcher i.e, link each researcher to a set of interests. In short, the system creates a bipartite graph with links between two disjoint sets: researchers and interests. We quantiatively analyze the consistency of this bipartite graph with respect to domain. We analyze this graph qualitatively by creating a network of blocks/communities of researchers based on their mutual interests. Details given below.

3.1. Entities and Relations

We have a disjoint list of entities: (a) researchers and (b) research interests. Our goal is to infer a list of interests for each researcher. For each researcher that was co-mentioned with a research interest, we create a link between researcher and that interest. For example, in a given dialog session with a user, researcher names mentioned in *FamousPeople* strategy are linked to interests mentioned in *Buzzwords* strategy. We repeat this process for researcher names and interests mentioned with respect to a talk i.e., *Keywords* associated with a particular talk are linked to *People* mentioned with that talk. This process produces a bipartite graph with connections between researchers and research interests. An example of researcher and predicted interests:

> **rich stern**: deep neural networks, speech recognition, signal processing, neural networks, machine learning, speech synthesis

We have 200 researchers (including the ones listed on the talk description), each mapped to a subset of interests from 485 unique interests. On average a researcher has 7.8 interests, with a standard deviation of 7.6 (this is because some researchers got more mentions across talks than others). We observed that some interests are linked to researchers more often than others — *machine learning*, *information retrieval* and *big data* are top-3 interests, linked with 49% of the researchers. To assess

Table 3: Mean Precision for 200 researchers breakdown by the "source" strategy used to acquire their name.

Source-of-Instance	Researchers	Mean Precision
Query Driven	21	86.2%
Egocentric	77	93.6%
Show & Ask	76	86.9%
Talk Description	61	89.5%
Overall	200	90.5%

the quality of predicted interests, we asked two senior Carnegie Mellon graduate students to label whether a predicted interest of a researcher is accurate. Table 3 shows the mean precision [5] for the predicted interests with a breakdown by source of researcher name and has good accuracy irrespective of the source.

To better understand how researchers are linked to interests and in general how researchers are aligned to each other, we construct an adjacency matrix of researchers. The details are described in the next subsection.

3.2. Analysis of Entity Network

To create an adjacency matrix of entities, i.e., the researchers, we compute the Jaccard index [25] for each pair of researchers based on their interests. The Jaccard index compares similarity of sample sets (say A and B) and is computed as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

We use the Jaccard index of two researchers as the weight of their connecting edge. We then convert the adjacency matrix to a network (an undirected graph) using a graph tool package[6]. To find communities or blocks in the resultant network, we use a stochastic block inference algorithm [26]. This algorithm tries to minimize description length (MDL) of the network (measured in nats-per-edge) to produce a block-partitioned version of the network. Intuitively, a block represent a set of nodes that more

---

[5]We use only precision because we do not have exhaustive list of relevant interests to measure the recall

[6]http://graph-tool.skewed.de

often interact within each other than with rest of the network; in our case, blocks are research communities. An illustration of a condensed version[7] of the network is shown in Figure 1 with research interests associated with each block shown in Figure 2.

Since the block inference algorithm optimizes the MDL, it may overfit the number of blocks needed to represent the network. Although in certain graphs (e.g., citation network) it is impossible to have modular or well-separated blocks, it is desirable to have reasonably separated blocks. Newman's modularity [27] is a metric, typically used to measure the strength of division in a graph [8]. To achieve a reasonable separation, we ran several iterations of the block inference algorithm, varying the minimum number of blocks required for the network. In Figure 3, we see that MDL is lowest (8.8) and MODULARITY is highest (0.08) when we set the minimum number of blocks to 16. This yielded a network of 17 blocks, as shown in Figure 1.

Figure 3: MDL and MODULARITY of network against min. number of blocks at initialization
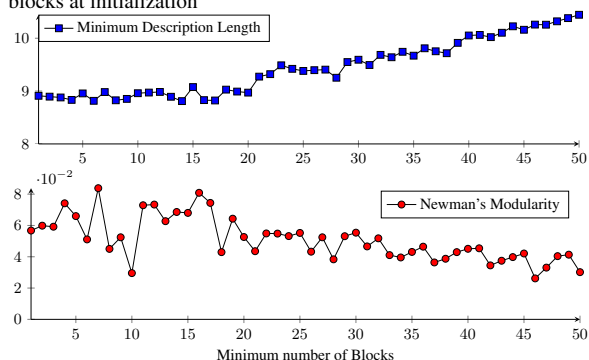
4. Using Knowledge: Query Expansion

Now, we want to know *whether the acquired knowledge is useful to the system* This can be demonstrated by showing that the EventSpeak System can respond more precisely to user queries by returning more relevant talks, after knowledge acquisition than before. Previous work [28] has shown that query expansion based on semantic networks can improve retrieval performance. We compare performance by expanding user query with acquired knowledge against the unexpanded queries.

4.1. Query Expansion Setup

We built a database of 160 research talks using abstracts, titles and speaker names as indices. For this purpose, we use Latent Semantic Indexing method as implemented in the GENSIM toolkit [29]. In the user study described earlier, we collected 40 queries (29 unique) about research areas and 40 (21 unique) about researchers — a total of 50 unique queries. We use these 50 queries to evaluate retrieval performance. Research area queries were expanded with their top-3 co-occurring research interests in the corpus. For researcher queries, we used the top-3 researcher's interests. Ten results with and without query expansion were retrieved. We then asked human judges (senior graduate students) to assess the relevance of each result on a scale of 1-4 (higher the better) with respect to a query. Our hypothesis is that results based on query expansion will have

higher relevance compared to the results without query expansion.

4.2. Query Expansion Results

We asked 5 human judges to rate the relevance of results from 50 queries with and without query expansion. We measured inter-annotator agreement by having a pair of judges annotate the same set of 20 queries. Overall we obtained ratings for 100 (5x10x2) queries: 5 sets of 10 queries each, rated by 2 judges per set — with good agreement (Cohen's $\kappa = 0.41$).

Table 4 shows the mean relevance, as rated by the annotators, for the retrieved talks with respect to the query. The query based expansion system outperforms the baseline, as observed in previous work [28]. Expansion works particularly well for "researcher" queries. One reason may be that person names may not have appeared in the talk description (and not indexed), but the research interests used to expand the query may appear in the talk descriptions (and indexed). We show that using this knowledge in expanding user queries can result in significantly[9] ($p < 0.01$) more relevant results (2.5/4 vs 1.8/4) than before acquisition.

Table 4: Mean relevance-per-query on scale of 1-4 (higher the better). Knowledge-based query expansion results are statistically ($p < 0.01$) more relevant than those without expansion.

QueryType	Without Expansion	With Expansion
Researcher	1.1 (stdev=0.8)	2.4 (stdev=0.6)
ResearchArea	2.2 (stdev=0.6)	2.5 (stdev=0.6)
Overall	1.8 (stdev=0.9)	**2.5** (stdev=0.6)

5. Conclusion and Future Work

We describe an agent that learns domain knowledge through dialog with users and uses it to build a semantic representation of an academic field. The system uses a set of strategies to collect entities (researchers and research interests). These entities are linked by their co-occurrence to produce a bipartite graph linking researchers and research interests.

To verify that the acquired knowledge is consistent, we asked human annotators to judge whether the interests predicted by the system were accurate. We found that the predicted interests for researchers have a high mean precision of 90.5%, i.e., annotators agree with the system's predictions in most cases. To analyze this knowledge qualitatively, we build a network of researchers connected through their mutual interests and divide this network into blocks using a block inference algorithm. This results in a set of blocks/communities of researchers (like a citation network) that covers the original academic field. We found that acquired knowledge used for query expansion provides more relevant results (2.5/4 vs 1.8/4), according to human judges, than without this acquired knowledge.

Knowledge acquisition involves cost (user's time) and incentive (new knowledge for system). Designing a policy to assess cost and incentive of acquisition, will be an immediate goal for this work. This policy depends on number of factors: identify the worth of unknown information, judge user's expertise about unknown information, etc. A policy with such factors may guide the system when to learn and what to learn.

---

[7]full-blown network is attached in the submission

[8]Modularity cannot capture blocks/communities in smaller graphs/networks, hence we do not use it directly for block partition

---

[9]using unpaired t-test

6. References

[1] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. H. Jr., and T. M. Mitchell, "Toward an Architecture for Never-Ending Language Learning," *Artificial Intelligence*, vol. 2, no. 4, pp. 1306–1313, 2010.

[2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," *SIGMOD 08 Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 1247–1249, 2008.

[3] N. Cristianini, J. Shawe-Taylor, and H. Lodhi, "Latent semantic kernels," *Journal of Intelligent Information Systems*, vol. 18, no. 2, pp. 127–152, 2002.

[4] P. Wang and C. Domeniconi, "Building semantic kernels for text classification using wikipedia," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 713–721.

[5] A. Moschitti, "Syntactic and semantic kernels for short text pair categorization," in *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 2009, pp. 576–584.

[6] A. Pappu and A. Rudnicky, "Predicting tasks in goal-oriented spoken dialog systems using semantic knowledge bases," in *Proceedings of the SIGDIAL 2013 Conference*. ACL, 2013, pp. 242–250.

[7] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *Journal of information science*, vol. 32, no. 2, pp. 198–208, 2006.

[8] "Google now." [Online]. Available: http://www.google.com/landing/now/

[9] "Apple siri." [Online]. Available: http://www.apple.com/ios/siri/

[10] L. Von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, 2006.

[11] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White *et al.*, "Vizwiz: nearly real-time answers to visual questions," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 2010, pp. 333–342.

[12] T. Spexard, S. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, and B. Krose, "BIRON, where are you? Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization," *Integration The Vlsi Journal*, no. section II, pp. 934–940, 2006.

[13] I. Lütkebohle, J. Peltason, L. Schillingmann, C. Elbrechter, B. Wrede, S. Wachsmuth, and R. Haschke, "The Curious Robot - Structuring Interactive Robot Learning," in *International Conference on Robotics and Automation*, ser. ICRA'09, IEEE. IEEE, 2009, pp. 4156–4162.

[14] H. Holzapfel, D. Neubig, and A. Waibel, "A dialogue approach to learning object descriptions and semantic categories," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 1004–1013, Nov. 2008.

[15] S. Lauria, G. Bugmann, T. Kyriacou, J. Bos, and A. Klein, "Training personal robots using natural language instruction," *Intelligent Systems, IEEE*, vol. 16, no. 5, pp. 38–45, 2001.

[16] A. I. Rudnicky, A. Pappu, P. Li, and M. Marge, "Instruction Taking in the TeamTalk System," in *Proceedings of the AAAI Fall Symposium on Dialog with Robots*, no. Dm, 2010, pp. 173–174.

[17] D. Bohus and A. I. Rudnicky, "Sorry, i didn't catch that!-an investigation of non-understanding errors and recovery strategies," in *Proceedings of the SIGDIAL*. ACL, 2005.

[18] G. Chung, S. Seneff, and C. Wang, "Automatic acquisition of names using speak and spell mode in spoken dialogue systems," in *Proceedings of the 2003 Conference of the iNorth American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 32–39.

[19] E. Filisko and S. Seneff, "Developing city name acquisition strategies in spoken dialogue systems via user simulation," in *6th SIGdial Workshop on Discourse and Dialogue*, 2005.

[20] R. Prasad, R. Kumar, S. Ananthakrishnan, W. Chen, S. Hewavitharana, M. Roy, F. Choi, A. Challenner, E. Kan, A. Neelakantan *et al.*, "Active error detection and resolution for speech-to-speech translation," in *Proceedings of IWSLT*, 2012.

[21] S. Stoyanchev, A. Liu, and J. Hirschberg, "Modelling human clarification strategies," in *Proceedings of the SIGDIAL*. ACL, 2013, pp. 137–141.

[22] N. Haas and G. Hendrix, *An approach to acquiring and applying knowledge*. Defense Technical Information Center, 1980.

[23] M. Witbrock, D. Baxter, J. Curtis, D. Schneider, R. Kahlert, P. Miraglia, P. Wagner, K. Panton, G. Matthews, and A. Vizedom, "An interactive dialogue system for knowledge acquisition in cyc," in *Proceedings of the 18th IJCAI*, 2003, pp. 138–145.

[24] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu, "Open mind common sense: Knowledge acquisition from the general public," in *CoopIS, DOA, and ODBASE*. Springer, 2002, pp. 1223–1237.

[25] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.

[26] T. P. Peixoto, "Parsimonious module inference in large networks," *Physical Review Letters*, vol. 110, no. 14, p. 148701, 2013.

[27] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[28] R. Navigli and P. Velardi, "An analysis of ontology-based query expansion strategies," in *Proceedings of the 14th European Conference on Machine Learning, Workshop on Adaptive Text Extraction and Mining, Cavtat-Dubrovnik, Croatia*, 2003, pp. 42–49.

[29] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, 2010, pp. 45–50.