

CTPs: Contextual Temporal Profiles for Time Scoping Facts using State Change Detection

Derry Tanti Wijaya
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
dwi@cmu.edu

Ndapandula Nakashole
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
ndapa@cmu.edu

Tom M. Mitchell
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
tom.mitchell@cmu.edu

Abstract

Temporal scope adds a time dimension to facts in Knowledge Bases (KBs). These time scopes specify the time periods when a given fact was valid in real life. Without temporal scope, many facts are under-specified, reducing the usefulness of the data for upper level applications such as Question Answering. Existing methods for temporal scope *inference* and *extraction* still suffer from low accuracy. In this paper, we present a new method that leverages temporal profiles augmented with context— Contextual Temporal Profiles (CTPs) of entities. Through change patterns in an entity’s CTP, we model the entity’s state change brought about by real world events that happen to the entity (e.g, hired, fired, divorced, etc.). This leads to a new formulation of the temporal scoping problem as a *state change detection problem*. Our experiments show that this formulation of the problem, and the resulting solution are highly effective for inferring temporal scope of facts.

1 Introduction

Recent years have seen the emergence of large Knowledge Bases (KBs) of facts (Carlson 2010; Auer 2007; Bollacker 2008; Suchanek 2007). While the wealth of accumulated facts is huge, most KBs are still sparsely populated in terms of temporal scope. Time information is an important dimension in KBs because knowledge is not static, it changes over time: people get divorced; countries elect new leaders; and athletes change teams. This means that facts are not always indefinitely true. Therefore, temporal scope has crucial implications for KB accuracy.

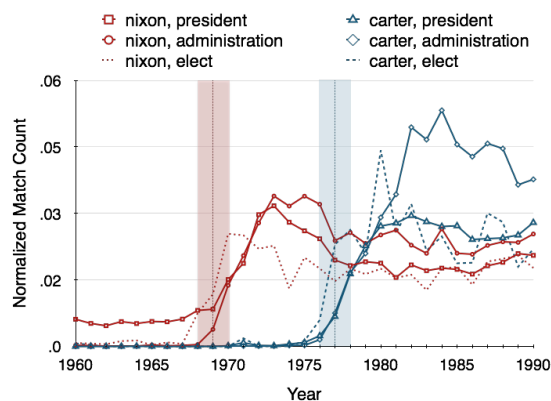


Figure 1: Behavior patterns of context uni-grams for the US presidency state change as seen in the Google Books N-grams corpus: the rise of ‘elect’, immediately followed by the rise of ‘administration’ and ‘president’.

Towards bridging the time gap in KBs, we propose a new method for temporal scope inference. Our method is based on leveraging aggregate statistics from a time-stamped corpus. First we generate Contextual Temporal Profiles (CTPs) of entities from contexts surrounding mentions of these entities in the corpus. We then detect change patterns in the CTPs. We then use these changes to determine when a given entity undergoes a specific state change caused by real world events. Our main insight is as follows: events that happen to an entity change the entity’s state and therefore its facts. Thus by learning *when* a given entity undergoes a specific state change, we can directly infer the time scopes of its facts. For example, in the *divorce event*, the person’s state changes from ‘married’ to ‘divorced’ hence the *hasSpouse* relation no longer applies to it, signaling the *end* time of its current *hasSpouse* value. In a *country election event*, the country’s state changes and it obtains a new value for its *hasPresident* relation.

Our method involves learning context units (uni-grams and bi-grams surrounding mentions of an entity) that are relevant to a given state change. For this we use a few seed examples of entities that have gone through the state change. For example, for the US presidency state change denoting the beginning of a US presidency, given seed examples such as (*Richard Nixon, 1969*) and (*Jimmy Carter, 1977*), relevant context units include uni-grams such as ‘administration’ and ‘elect’, which are common to both CTPs in 1969 and 1977 respectively. Secondly, we learn the mention behavior of these context units for an entity undergoing a given state change (section 3 has more details). Figure 1 shows a motivating example, we see the behavior patterns of context uni-grams for the US presidency state change: the rise of ‘elect’ at the beginning of presidencies, immediately followed by the rise of ‘administration’ and ‘president’ in the context of the entities, *Nixon* and *Carter*.

2 Related work

Prior work mainly falls into two categories: i) methods that extract temporal scope from text, at the time of fact extraction; ii) methods that infer temporal scope from aggregate statistics in large Web corpora. Early methods mostly fall under category i); Timely YAGO (Wang 2010), TIE (Ling 2010), and PRAVDA (Wang 2011) are three such methods. Timely YAGO applies regular expressions to Wikipedia infoboxes to extract time scopes. It is therefore not applicable to any other corpora but Wikipedia. The TIE (Ling 2010) system produces a maximal set of events and their temporal relations based on the text of a given sentence. PRAVDA uses textual patterns along with a graph-based re-ranking method. Methods falling under category i) have the downside that it is unclear how they can be applied to facts that are already in the knowledge base. Only one other approach learned time scopes from aggregate corpus statistics, a recent system called CoTS (Talukdar 2012b). CoTS uses temporal profiles of facts and how the mentions of such facts rise and fall over time. However, CoTS is based on frequency counts of fact mentions and does not take into account state change inducing context. For example, to find the time scope of Nixon presidency, CoTS uses the rise and fall of the mention ‘nixon’ and ‘president’ over time. To improve accuracy,

CoTS combined this frequency signal with manually supplied constraints such as the functionality of the US presidency relation to scope the beginning and end of Nixon presidency. In contrast, the proposed system does not require constraints as input.

There have also been tools and competitions developed to facilitate temporal scope extraction. TARSQI (Verhagen 2005) is a tool for automatically annotating time expressions in text. The TempEval (Verhagen 2007) challenge has led to a number of works on temporal relation extraction (Puscasu 2007; Yoshikawa 2009; Bethard 2007).

3 Method

Given an entity and its Contextual Temporal Profile (CTP), we can learn *when* such an entity undergoes a specific state change. We can then directly infer the *begin* or *end* time of the fact associated with the state change.

The CTP of an entity at a given time point t contains the context within which the entity is mentioned at that time. Our method is based on two related insights: i) the context of the entity at time t reflects the events happening to the entity and the state of the entity at time t . ii) the *difference in context* before, at time $t - 1$, and after, at time t , reflect the associated state change at time t . However an entity can undergo a multiplicity of changes at the same time. Thus both the contexts and the differences in contexts can contain information pertaining to several state changes. We therefore need a way of determining which part of the context is relevant to a given state change sc_i . To this end, we generate what we refer to as an *aggregate* state vector, $Vs(\bar{e}, sc_i)$ for a hypothetical average entity \bar{e} undergoing state change sc_i . We generate $Vs(\bar{e}, sc_i)$ from the CTPs of a seed set of entities at the time they undergo state change sc_i .

3.1 Learning State and State Change Vectors

To build CTPs for entities, we use two time-stamped corpora: the Google Books Ngram corpus (Michel 2011); and the English Gigaword (Graff 2003) corpus. The Google Books Ngram corpus contains n-grams for $n = 1 - 5$; along with occurrence statistics from over about 5 million digitized books. The English Gigaword (Graff

2003) corpus contains newswire text from 1994-2008. From these corpora, we use the time granularity of a year as it is the finest granularity common to both corpora.

Definition 1 (Contextual Temporal Profile)

The Contextual Temporal Profile (CTP) of an entity e at time t , $C_e(t)$, consists of the context within which e is mentioned. Specifically $C_e(t)$ consists of uni-grams and bi-grams generated from the 5-grams(Google Books Ngram) or sentences (Gigaword) that mention e at time t .

Notice that the CTPs can contain context units (bi-grams or uni-grams) that are simply noise. To filter the noise, we compute *tf-idf* statistics for each contextual unit and only retain the top k ranking units in $C_e(t)$. In our experiments, we used $k = 100$. We compute *tf-idf* by treating each time unit t as a document containing words that occur in the context of e (Wijaya 2011).

Furthermore, CTPs may contain context units attributed to several state changes. We therefore tease apart the CTPs to isolate contexts specific to a given state change. For this, our method takes as input a small set of seed entities, $\mathcal{S}(sc_i)$, for each type of state change. Thus for the US presidency state change that denotes the beginning of a US presidency, we would have seeds as follows: (*Richard Nixon, 1969*), (*Jimmy Carter, 1977*). From the CTPs of the seeds for state change sc_i , we generate an aggregate state vector, $Vs(\bar{e}, sc_i)$. To obtain the few dozen seeds required by our method, one can leverage semi-structured sources such as Wikipedia *infoboxes*, where relations e.g., *spouse* often have time information.

Definition 2 (Aggregate State Vector for \bar{e})

The aggregate state vector of a mean entity \bar{e} for state change sc_i , $Vs(\bar{e}, sc_i)$, is made up of the contextual units from the CTPs of entities in the seed set $\mathcal{S}(sc_i)$ that undergo state change sc_i . Thus, we have: $Vs(\bar{e}, sc_i) = \frac{1}{|\mathcal{S}(sc_i)|} \sum_{e,t:(e,t) \in \mathcal{S}(sc_i)} C_e(t)$.

Thus, the state vector $Vs(\bar{e}, sc_i)$ reflects events happening to \bar{e} and the state of \bar{e} at the time it undergoes the state change sc_i . Additionally, we compute another type of aggregate vector, *aggregate change vector* $\Delta Vs(\bar{e}, sc_i)$ to capture the change patterns in the context units of \bar{e} . Recall that context units rise or fall due to state change, as seen earlier in Figure 1.

Definition 3 (Aggregate Change Vector for \bar{e})

The aggregate change vector of a mean entity \bar{e} for state change sc_i , $\Delta Vs(\bar{e}, sc_i)$, is made up of the change in the contextual units of the CTPs of entities in the seed set $\mathcal{S}(sc_i)$ that undergo state change sc_i . Thus, we have: $\Delta Vs(\bar{e}, sc_i) = \frac{1}{|\mathcal{S}(sc_i)|} \sum_{e,t:(e,t) \in \mathcal{S}(sc_i)} C_e(t) - C_e(t - 1)$.

The aggregate state vector $Vs(\bar{e}, sc_i)$ and the aggregate change vector $\Delta Vs(\bar{e}, sc_i)$ are then used to detect state changes.

3.2 Detecting State Changes

To detect state changes in a previously unseen entity e_{new} , we generate its state vector, $C_{e_{new}}(t)$, and its change vector, $\Delta C_{e_{new}}(t) = C_{e_{new}}(t) - C_{e_{new}}(t - 1)$, for every time point t . We consider every time point t in the CTP of the new entity to be a candidate for a given state change sc_i , which we seek to determine whether e_{new} goes through and at which time point. We then compare the state vector and change vector of every candidate time point t to the aggregate state and aggregate change vector of state change sc_i . We use cosine similarity to measure similarities between the state vector and the aggregate state vector and between the change vector and the aggregate change vector. To combine these two vector similarities, we sum the state vector and change vector similarities. In future we can explore cross validation and a separate development set to define a weighted sum for combining these two similarities.

The highest ranking candidate time point (most similar to the aggregate state and aggregate change vector) is then considered to be the start of state change sc_i for the new entity e_{new} .

4 Experiments

We carried out experiments to answer the following questions: *Is treating temporal scoping as state change detection in Contextual Temporal Profiles(CTPs) effective? Do CTPs help improve temporal scope extraction over context-unaware temporal profiles?*

4.1 Methods under Comparison

We answer these questions by comparing to the following methods.

1. **CoTS** a state-of-the-art temporal scoping system (Talukdar 2012b)

2. **MaxEnt** a baseline to which CoTS was compared. It is a Maximum Entropy classifier trained separately for each relation using normalized counts and gradients of facts as features. An Integer Linear Program (ILP) is used to predict which facts are active at which times. This is done based on the output of the MAXENT classifier together with temporal intra-relation constraints that regulate the temporal scoping of one or more facts from a *single* relation (e.g., FUNCTIONAL constraints on US President relation that regulate that at most one fact from the relation can be true at any given time i.e., there is only one US President at any given time).

3. MaxEnt + Intra Relation Constraints

MaxEnt with cross relation constraints added: constraints that couple facts from *multiple* relations e.g., a constraint that *Al Gore's* vice presidency is aligned exactly with *Bill Clinton's* presidency.

We evaluate on the same set of facts as CoTS and its baselines: facts from the US Administration domain (US President, US Vice President, and US Secretary of State); and facts from the Academy Awards domain (Best Director and Best Picture). The number of facts per relation are as follows: US President, 9; US Vice President, 12; US Secretary of State, 13; Best Director, 14; and Best Picture, 14. Our method however is not specific to these relations from these two domains. Since our method does not depend on temporal constraints, the method can work a very different domain, for example one where many facts can exist for any time span without being superseded by another, as long as the entities involved experience a change of state. Thus, it can be applied to relations like *spouse*, even though many people are married in a year as these people change state from *single* or *engaged* to *married*.

Similar to CoTS, the datasets from which the CTPs were generated are as follows: The Google Books Ngram (1960-2008) dataset (Michel 2011) for the US Administration domain and the English Gigaword (1994-2008) dataset (Graff 2003) for Academy Award domain.

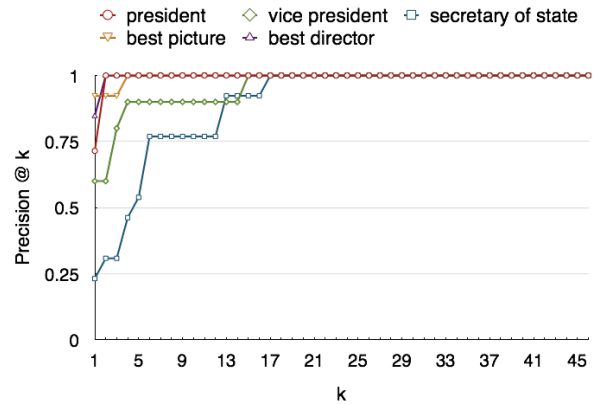


Figure 2: Precision @ k using Contextual Temporal Profiles.

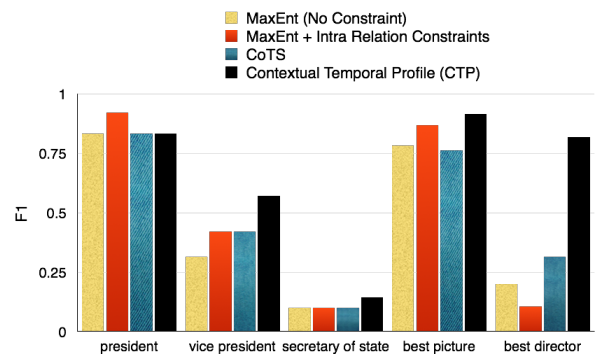


Figure 3: Comparison of F1 scores with CoTS and other baselines.

4.2 CTPs Begin time precision

To compute precision we used cross validation, in particular, leave-one-out cross validation due to the small number of facts per relation. We predict the *begin* time of each fact, the time the fact starts to be valid. True begin times were determined by a human annotator. This human annotated data formed the gold-standard which we used to determine Precision (P), Recall (R), and the F1 measure. All evaluations were performed at the year level, the finest granularity common to the two time-stamped datasets.

For our first experiment, we report the average precision@k, where $k=1$ to n , where $n=47$ is the number of years between 1960 to 2008 to select from. As can be seen in Figure 2, precision quickly reaches 1 for most relations. The true begin time is usually found within top $k=5$ results.

4.3 Comparison to baselines

For our second experiment, we compared to the F1 scores of CoTS and other baselines in (Talukdar 2012b). As can be seen in Figure 3, our CTPs approach gives comparable or better F1 (@ $k=1$) than systems that use only plain temporal profiles, even when these systems are supplemented with many carefully crafted, hand-specified constraints.

We note that the performance on the *US Secretary of State* relation is low in both CoTS (Talukdar 2012b) and in our approach. We found that this was due to few documents mentioning the “secretary of state” in Google Books Ngram dataset. This leads to weak signals for predicting the temporal scope of secretary of state appointments.

We also observe that the uni-grams and bi-grams in the train CTPs and change vectors reflect meaningful events and state changes happening to the entities (Table 1). For example, after ‘becoming president’ and ‘taking office’, US presidents often see a drop in mentions of their previous (job title state) such as ‘senator’, ‘governor’ or ‘vice president’ as they gain the ‘president’ state.

4.4 Discussion

Overall, our results show that our method is promising for detecting *begin* time of facts. In its current state, our method performs poorly on inferring *end* times as contexts relevant to a fact often still mentioned with the entity even after the fact ceases to be valid. For example, the entity *Al Gore* is still mentioned a lot with the bi-gram ‘vice president’ even after he is no longer a vice president. Prior work, CoTS, inferred *end* times by leveraging manually specified constraints, e.g., that there can only be one vice president at a time: the beginning of one signals the end of another (Talukdar 2012b). However such methods do not scale due to the amount of constraints that must be hand-specified. In future, we would like to investigate how to better detect the *end* times of facts.

5 Conclusion

This paper presented a new approach for inferring temporal scopes of facts. Our approach is to reformulate temporal scoping as a state change detection problem. To this end, we introduced Contextual Temporal Profiles (CTPs) which are entity temporal profiles enriched with relevant context.

| Relation | CTP State Context | Unigrams and Bigrams in CTP Change Vectors |
|--------------|---|--|
| US President | was elected, took office, became president | vice president (-), by president (+), administration (+), senator (-), governor (-), candidate president (-) |
| Best Picture | nominated for, to win, won the, was nominated | best picture (+), hour minute (-), academy award (+), oscar (+), nominated (+), won (+), star (-), best actress (+), best actor (+), best supporting (+) |

Table 1: Example behavior of various contextual units (unigrams and bigrams) automatically learned in the train CTPs and change vector. The (+) and (-) signs indicate rise and fall in mention frequency, respectively.

From the CTPs, we learned change vectors that reflect change patterns in context units of CTPs. Our experiments showed that the change patterns are highly relevant for detecting state change, which is an effective way of identifying *begin* times of facts. For future work, we would like to investigate how our method can be improved to do better at detecting fact *end* times. We also would like to investigate time-stamped corpora of finer-grained granularity such as day. This information can be obtained by subscribing to daily newsfeeds of specific entities.

Acknowledgments

We thank members of the NELL team at CMU for their helpful comments. This research was supported by DARPA under contract number FA8750-13-2-0005 and in part by Fulbright and Google Anita Borg Memorial Scholarship.

References

- A. Angel, N. Koudas, N. Sarkas, D. Srivastava: Dense Subgraph Maintenance under Streaming Edge Weight Updates for Real-time Story Identification. In *Proceedings of the VLDB Endowment*, PVLDB 5(10):574–585, 2012.
- S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z.G. Ives: DBpedia: A Nucleus for a Web of Open Data. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, pages 722–735, Busan, Korea, 2007.
- M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, O. Etzioni: Open Information Extraction from

- the Web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2670–2676, Hyderabad, India, 2007.
- S. Bethard and J.H. Martin. Cu-tmp: Temporal relation classification using syntactic and semantic features. In *SemEval-2007*, 2007.
- K. D. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor: Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages, 1247-1250, Vancouver, BC, Canada, 2008.
- A. Carlson, J. Betteridge, R.C. Wang, E.R. Hruschka, T.M. Mitchell: Coupled Semi-supervised Learning for Information Extraction. In *Proceedings of the Third International Conference on Web Search and Web Data Mining (WSDM)*, pages 101–110, New York, NY, USA, 2010.
- A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr., T. M. Mitchell: Toward an Architecture for Never-Ending Language Learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI) 2010*.
- L. Del Corro, R. Gemulla: ClausIE: clause-based open information extraction. In *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, pages 355-366. 2013.
- A. Das Sarma, A. Jain, C. Yu: Dynamic Relationship and Event Discovery. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining (WSDM)*, pages 207–216, Hong Kong, China, 2011.
- A. Fader, S. Soderland, O. Etzioni: Identifying Relations for Open Information Extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1535–1545, Edinburgh, UK, 2011.
- D. Graff, J. Kong, K. Chen, and K. Maeda. English gigaword. Linguistic Data Consortium, Philadelphia, 2003.
- C. Havasi, R. Speer, J. Alonso. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. In *Proceedings of the Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria, 2007.
- J. Hoffart, F. Suchanek, K. Berberich, E. Lewis-Kelham, G. de Melo, G. Weikum: YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. In *Proceedings of the 20th International Conference on World Wide Web (WWW)*, pages 229–232, Hyderabad, India. 2011.
- X. Ling and D.S. Weld. Temporal information extraction. In *Proceedings of AAAI*, 2010.
- Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Holberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden: Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176182.
- N. Nakashole, M. Theobald, G. Weikum: Scalable Knowledge Harvesting with High Precision and High Recall. In *Proceedings of the 4th International Conference on Web Search and Web Data Mining (WSDM)*, pages 227–326, Hong Kong, China, 2011.
- N. Nakashole, T. Tylenda, G. Weikum: Fine-grained Semantic Typing of Emerging Entities. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1488-1497, 2013.
- N. Nakashole, T. M. Mitchell: Language-Aware Truth Assessment of Fact Candidates In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1009-1019, 2014.
- G. Puscasu. Wvali: Temporal relation identification by syntactico-semantic analysis. In *Proceedings of the 4th International Workshop on SemEval*, 2007.
- J. Pustejovsky, J. Castano, R. Ingria, R. Sauri, R. Gaizauskas, A. Setzer, G. Katz, and D. Radev. Timeml: Robust specification of event and temporal expressions in text. In *Fifth International Workshop on Computational Semantics*, 2003.
- P. P. Talukdar, D. T. Wijaya, Tom M. Mitchell: Acquiring temporal constraints between relations. In *Proceeding of the 21st ACM International Conference on Information and Knowledge Management*, pages 992-1001, CIKM 2012.
- P. P. Talukdar, D. T. Wijaya, T. Mitchell: Coupled temporal scoping of relational facts. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012.
- M. Verhagen, I. Mani, R. Sauri, R. Knippen, S.B. Jang, J. Littman, A. Rumshisky, J. Phillips, and J. Pustejovsky. Automating temporal annotation with tarsqi. In *Proceedings of the ACL Session on Interactive poster and demonstration sessions*, 2005.
- M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, and J. Pustejovsky. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.
- D. T. Wijaya, and R. Yeniterzi: Understanding semantic change of words over centuries. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversity on the social web*. ACM, 2011.

- F. M. Suchanek, G. Kasneci, G. Weikum: Yago: a Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web (WWW)* pages, 697-706, Banff, Alberta, Canada, 2007.
- Y. Wang, M. Zhu, L. Qu, M. Spaniol, and G. Weikum: Timely yago: harvesting, querying, and visualizing temporal knowledge from wikipedia. In *Proceedings of the 13th International Conference on Extending-Database Technology*, 2010.
- W. Wu, H. Li, H. Wang, K. Zhu: Probase: A Probabilistic Taxonomy for Text Understanding. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 481–492, Scottsdale, AZ, USA, 2012.
- Y. Wang, B. Yang, L. Qu, M. Spaniol, and G. Weikum: Harvesting facts from textual web sources by constrained label propagation. In *Proceedings of CIKM*, 2011.
- K. Yoshikawa, S. Riedel, M. Asahara, and Y. Matsumoto. Jointly identifying temporal relations with markov logic. In *Proceedings of ACL*, 2009.