CHAPTER XX

Credibility assessment and inference for fusion of hard and soft information

R.C.Núñez[*], T.L.Wickramarathne[*], K.Premaratne[*], M.N.Murthi[*], S.Kübler[‡], M.Scheutz[◊], M.A.Pravia[+]

[*]University of Miami
Coral Gables, FL, USA
{nunez, t.wickramarathne}@umiami.edu, {kamal, mmurthi}@miami.edu

[‡]Indiana University
Bloomington, IN, USA
skuebler@indiana.edu

[◊]Tufts University
Medfort, MA, USA
mscheutz@cs.tufts.edu

[+]BAE Systems
Burlington, MA, USA
marco.pravia@baesystems.com

ABSTRACT

Effectively combining multiple (and complementary) sources of information is becoming one of the most promising paths for increased accuracy and more detailed analysis in numerous applications. Neuroscience, business analytics, military intelligence, and sociology are among the areas that could significantly benefit from properly processing diverse data sources. However, traditional methods for combining multiple sources of information are based on slow or impractical methods that rely either on vast amounts of manual processing or on suboptimal representations of data. We introduce an analytical framework that allows automatic and efficient processing of both hard (e.g., physics-based sensors) and soft (e.g., human-generated) information, leading to enhanced decision-making in multi-source environments. This framework combines Natural Language Processing (NLP) methods for extracting information from soft data sources and the Dempster-Shafer (DS) Theory of Evidence as the common language for data representation and inference. The steps in the NLP module consist of part-of-speech tagging, dependency parsing, coreference resolution, and a conversion to semantics based on first order logic representations. Compared to other methods for handling uncertainties, DS theory provides an environment that is better suited for capturing data models and imperfections that are common in soft data. We take advantage of the fact that computational complexity typically associated with DS-based methods is continually decreasing with both the availability of better processing systems, as well as with improved processing algorithms such as conditional approach to evidence updating/fusion. With an adequate environment for numerical modeling and processing, two additional elements become especially relevant, namely: (1) assessing source credibility, and (2) extracting meaning from available data. Regarding (1), it is clear that the lack of source credibility estimation (especially with human-generated information) could direct even the most powerful inference methods to the wrong conclusions. To address this issue we present consensus algorithms that mutually constrain the data provided by each of the sources to assess their individual credibility. This process can be reinforced to get improved results by incorporating (possibly partial) information from physical sensors to validate soft data. At the end of a credibility estimation process, every piece of information can be properly scaled prior to any inference process. Then, meaning extraction (i.e., (2)) becomes possible by applying the desired inference method. Special consideration must be taken to ensure that the selected inference method preserves the quality and accuracy of the original data as much as possible, as well as the relations among different sources of information and among data. To accomplish this, we propose using first-order logic (FOL) in the DS theoretic framework. Under this approach, soft information (in the form of natural language) is analyzed syntactically and for discourse structure, and consequently converted into FOL statements representing the semantics. Processing of these statements through an "uncertain logic" DS methodology renders bodies of evidence (BoE) that, combined with experts' opinions stored in knowledge bases, can be fused to provide accurate solutions to a wide variety of queries. Examples of queries include finding or refining groups of suspects in a crime scene, validating credibility of witnesses, and categorizing data in the web. When hard-sensor data is also incorporated in the inference process, challenging applications such as multi-source detection, tracking, and intent detection, could also be addressed with the proposed solution.

Keywords: Evidence Fusion, Source Credibility, Credibility Estimation, Consensus, Belief Theory, Theory of Evidence, First Order Logic, Meaning Extraction, Natural Language Processing, Dependency Parsing, Coreference Resolution.

1 INTRODUCTION

The development of new sensing and data acquisition technologies is occurring at such a fast pace that it is triggering a need for more sophisticated meaning extraction and inference methods. These inference schemes need to take advantage of the increased amounts of information, producing more accurate and complete solutions to varied problems. This information, in general, can be classified as "hard" or "soft". "Hard information" refers to information generated by physics-based sources, and "soft information" refers to information generated from human-based sources, including human reports, text and audio communications, and open sources such as newspapers, radio/TV broadcasts, and web sites.

Solutions for meaning extraction and inference have typically targeted either hard information (e.g., sensor networks) or soft information (e.g., data mining). However, simultaneously using both hard and soft information is still mostly a human-intensive task, with very little research addressing this hard/soft information fusion application (Pravia, et. al., 2008).

Aimed at addressing this issue, we introduce a general model for automated analysis of hard and soft information. As an application of this technology, consider the following scenario.[1] A team of experts is trying to assess the credibility of witnesses of a crime scene. The messages provided by the witnesses, as they were documented, are:

Witness 1 (W1): "The suspect was driving a black SUV";
Witness 2 (W2): "The suspect was driving a white sedan";
Witness 3 (W3): "The suspect was driving a white vehicle".

Each of the witnesses was asked to rate, from 0 to 100, how certain he/she was on the information they provided. They answered 80%, 90%, and 95%, for W1, W2, and W3, respectively. In addition to the information provided by witnesses, the team of experts has access to video surveillance (VS) reports that identified the suspect's vehicle as a light-colored sedan. This type of report has been characterized as being 98% accurate.

Having this information, is it possible to estimate the credibility of the witnesses? Is it possible to refine the crime-scene scenario? Although this simple scenario can be easily solved by humans, our work aims at defining a framework that allows automatic analysis of this type of events, especially when hundreds of thousands of pieces of information are available, all of them potentially providing valuable information for the human experts.

A general framework for solving this type of problem is shown in Figure 1. In this framework, soft data is converted into first-order logic (FOL) constructs by a Natural Language Processing (NLP) module. FOL is preferred for the semantic representation because it preserves a higher amount of information compared to other methods (e.g., RDF Graphs). A combination of semantic representation

---

[1] We introduce this scenario as a running example that will allow us to easily describe each step of our hard and soft information fusion process.

methods with higher and lower levels of detail could also be used as means of reducing complexity. These logic constructs are quantified and mathematically modeled (e.g., via probability or belief functions), and their credibility is assessed. An alternative for assessing credibility is based on finding consensus among information sources. Distance to consensus can be used as a measure of credibility. With such a measure, the data could be properly weighted for further processing in meaning extraction and inference.

In the remainder, we introduce the methods that we have designed for each of the components in our hard/soft information fusion framework, and, as an illustration of the techniques, we apply them for processing the scenario described above.

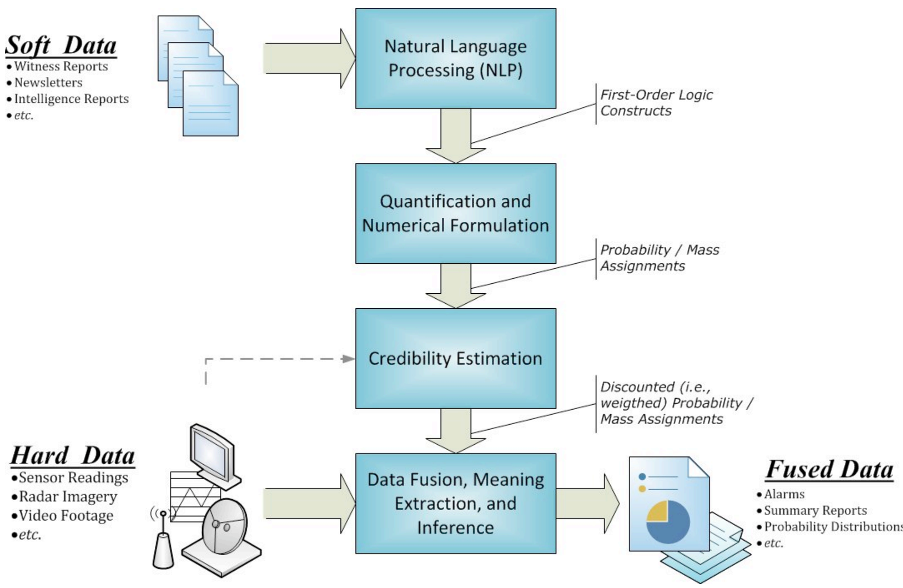

Figure 1 Fusion of hard and soft information.

2 BUILDING NUMERICAL MODELS FROM TEXTUAL DATA

2.1 Natural Language Processing

The goal of the NLP module is to accept plain text, such as witness statements, interrogation protocols, texts from the WWW, or any other textual source, and produce an analysis that allows further processing in the Dempster-Shafer (DS) framework. Overall, the NLP analysis faces two problems: (1) efficiency concerns and (2) out-of-domain data. With regard to efficiency concerns, syntactic parsing and coreference resolution are computationally very intensive steps. Additionally, a traditional NLP architecture is built as a pipeline, in which texts are processed sentence by sentence, i.e. one sentence is analyzed by the first module, then sent to the second, etc. We approach this problem by using a fully incremental architecture, in which words are processed as they come in, and partial results are directly passed to the next module without waiting for the full sentence to be analyzed. This necessitates changes in the individual modules, which cannot rely on having access to context beyond the word currently processed. Problem (2) refers to the fact that all NLP modules are based on supervised machine learning, and thus need to be trained. The only available training set is often a part of the Wall Street Journal, which has been annotated on different levels (cf. e.g., Marcus et al., 1993). For this reason, we need to develop methods that allow us to adapt the learned model to the domain of texts that need to be analyzed (Kübler et al., 2010, Kübler and Baucom, 2011).

The first step in the NLP module is part-of-speech (POS) tagging, which assigns words classes to words in a sentence. Thus witness statement W1 would be assigned the following parts of speech: The/DT suspect/NN was/VBD driving/VBG a/DT black/JJ SUV/NNP. 'The' and 'a' are articles, 'suspect' is a noun, 'was' is a verb in past tense, 'driving' a present participle, 'black' is an adjective, and 'SUV' is a name. We use an n-best, anytime implementation of a Markov model tagger.

The POS tagged sentences serve as input for a dependency parser. Dependency parsing performs a syntactic analysis. We use a dependency parser, which assigns dependency relations to pairs of words in a sentence. The dependency analysis for W1 is shown in Figure 2. The analysis shows that 'suspect' is the subject of 'was', 'driving' is a verbal complement of 'was', 'SUV' is the direct object of 'driving', and the two articles modify the nouns. As parser, we use MINK (Cantrell, 2009), a fully incremental implementation of MaltParser (Nivre et al., 2007).

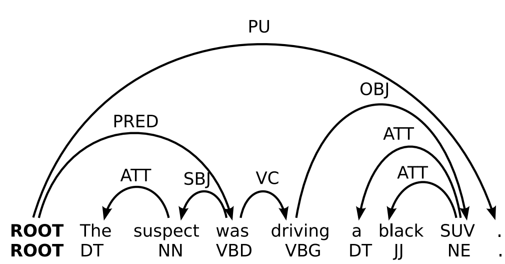

Figure 2 The dependency analysis of W1

After the syntactic analysis, we perform coreference resolution. In this step, we determine which expressions in the document refer to the same entity. That is, if we had a fourth witness statement:

Witness 4 (W4): "I saw the suspect driving a black SUV; he parked the vehicle right in front of the shop",

then coreference resolution will tell us that 'the suspect' and 'he' refer to the same person and that 'a black SUV' and 'the vehicle' refer to the same car. For coreference resolution, we use UBIU, a robust, multilingual system using memory-based learning for classifying pairs of potentially coreferent mentions (Zhekova and Kübler, 2010, Zhekova and Kübler, 2011). For the future, we are planning on extending the system to cross-document coreference resolution, which will enable us to cross-link mentions of persons across different documents, thus giving us sets of sentences that can be used for DS inference.

After coreference resolution, we convert the sentence into a semantic representation based on first order logic statements. Thus, W1 in our running example (see Section 1 above), would be represented as ∃x, y: WasDriving(x, y) ^ Suspect(x) ^ Black(y) ^ Suv(y), stating that there is an entity x who is a suspect, and an entity y which is an SUV, and that x was driving y. The conversion is lexicon-based (Kübler et al., 2011).

2.2 Basic Probability Assignments for FOL Constructs

With a set of FOL sentences available, the next problem that must be tackled is that of quantifying this information in such a way that uncertainty in the information is properly modeled, and that the inference engine can use it.

Although FOL has traditionally been one of the preferred modeling frameworks for resolution and inference, FOL is not designed for handling problems with uncertainties. Methods such as uncertain reasoning and probabilistic logic extend FOL for solving these problems (Genesereth and Nilsson, 1987). These methods improve the scope of problems that can be solved based on FOL, but they do not address the cases of incomplete data, and they cannot be used for easily describing information uncertainty by lower and upper probability bounds. We propose modeling data and making inference using Dempster-Shafer (DS) theory (Shafer, 1976) for addressing these issues. DS theory has been successfully used in applications such as rule mining (Hewawasam, et. al., 2007) and target identification (Ristic and Smets, 2005).

Current approaches for building DS models for logic operators provide models for particular uncertain operators in propositional logic. For example, Benavoli, et. al., (2008) introduce a method for modeling uncertain implication rules using DS models. We are enhancing model-building strategies by defining a method that incorporates logic quantifiers (hence, FOL models), and preserves fundamental logic properties (e.g., associativity, commutativity, distributivity, and idempotency) for a set of basic logic operators (i.e., `not, and, or`, and implication rules).

Based on our FOL models for DS, the witness sentences in our example can be converted into probability (or mass) assignments. Let us assume that the output of the NLP stage for our running example is the following:

W1: `∃x,y: WasDriving(x, y) ^ Suspect(x) ^ Black(y) ^ Suv(y);`

W2: `∃x,y: WasDriving(x,y) ^ Suspect(x) ^ White(y) ^ Sedan(y);`

W3: `∃x,y: WasDriving(x, y) ^ Suspect(x) ^ White(y) ^ Vehicle(y);`

where $x \in \Theta_{suspects}$, and $y \in \Theta_{vehicles}$. $\Theta_{suspects}$ and $\Theta_{vehicles}$ are called the frame of discernment (FoDs), i.e., the sets that define the groups of elementary events related to the problem.

For a decision process, we typically need to incorporate some domain knowledge. In this case we assume:

$\Theta_{vehicles} = \{ \Theta_{color} \times \Theta_{type} \}$;
$\Theta_{color} = \{$ white, silver, red, brown, black $\}$;
$\Theta_{light} = \{$ white, silver $\} \subset \Theta_{color}$;
$\Theta_{dark} = \{$ brown, black $\} \subset \Theta_{color}$;
$\Theta_{type} = \{$ sedan, jeep, SUV, truck $\}$.

DS theory models are defined by mass assignments. A mass assignment (or basic belief assignment) is a mapping $m_\Theta(\cdot)$: $2 \rightarrow [0, 1]$, such that $\sum_{A \subseteq \Theta} m_\Theta(A) = 1$ and $m_\Theta(\varnothing) = 0$. The mass assignment measures the support assigned to proposition $A \subseteq \Theta$. The triple $\{\Theta, \Im, m(\cdot)\}$, with $\Im$ being the set of all elements for which m(A) > 0, is called the body of evidence (BoE).

The mass assignments corresponding to witness statements and video surveillance reports are:

W1: $m_1(\{black\} \times \{SUV\}) = 0.80$; $m_1(\Theta_{vehicles}) = 0.20$;
W2: $m_2(\{white\} \times \{sedan\}) = 0.90$; $m_2(\Theta_{vehicles}) = 0.10$;
W3: $m_3(\{white\} \times \Theta_{type}) = 0.95$; $m_3(\Theta_{vehicles}) = 0.05$;
VS: $m_4(\{white, silver\} \times \{sedan\}) = 0.98$; $m_4(\Theta_{vehicles}) = 0.02$.

These mass assignments are obtained by using DS fusion based on the conditional update equation (CUE) (Premaratne, et. al., 2009). The fusion operators must be properly tuned for handling logic operations. The mass assignments can then be used for credibility assessment and inference, as is described next.

3 ESTIMATING CREDIBILITY OF SOURCES

It is very important, especially when dealing with multiple sources of information, to account for the credibility (i.e., trustworthiness) of the sources. In the DS framework, it is possible to account for this credibility by a procedure called discounting.

The issue becomes then, being able to estimate this credibility measure (in some applications such as judge or jury trials, the whole problem is precisely assessing the credibility of witnesses). When an adequate number of sources are considered, it is not unreasonable to assume that the truth is reflected in the majority opinion. If this majority opinion can be established via some rational aggregation procedure, the very aggregate, often referred to as a consensus, can in turn be used for credibility estimation.

We propose a consensus-based technique for credibility estimation of evidence in the absence of the ground truth. The credibility of a BoE E (i.e., a particular piece of evidence) can be defined as (Wickramarathne, et. al., 2012):

$$\mathrm{Cr}\,(E) = \left(1 - \mathrm{dist}(E, E^t)^\lambda\right)^{1/\lambda},$$

with E^t denoting a BoE that contains the ground truth, and $\lambda \in \Re^+$. This definition requires the computation of consensus BoEs. These BoEs can be computed using an iterative procedure based on the CUE. A detailed description of this credibility estimation technique as well as of the BoE update procedure can be found in (Wickramarathne, et. al., 2012).

Based on this methodology, the consensus BoE in our example is given by:

$$m_{consensus}(\ \{white\} \times \{sedan\}\ ) = 1.00,$$

and the distance to consensus as well as the credibility of the sources become:

dist(WS1, Consensus) = 0.9154; Cr (WS1) = 0.0846
dist(WS2, Consensus) = 0.0975; Cr (WS2) = 0.9025
dist(WS3, Consensus) = 0.8786; Cr (WS3) = 0.1214,

with the credibility estimated using $\lambda = 1$. This estimated credibility can then be used for evidence discounting prior to fusion operations in meaning extraction and inference.

4 MEANING EXTRACTION AND INFERENCE

As mentioned above, DS fusion offers several advantages over other meaning extraction methods, given that it incorporates a more rigorous modeling of uncertainties, and that it allows relevant fusion operations even in the presence of incomplete data. Nevertheless, there are some challenges typically associated with DS-based fusion: (1) limitations when handling information with dissimilar FoDs; (2) sensitivity to contradictory evidence; and (3) computational complexity.

The first of these challenges is particularly magnified when soft information is processed. When dealing with multiple sources of soft data, it is not uncommon to find data generated from non-identical FoDs. For example, the information contained in a public database of vehicles belonging to town residents would have a much larger, but not completely disjoint, scope than the vehicles that had been recorded at a checkpoint. Conventional DS methods are not suitable for such problems. Moreover, conventional DS fusion methods are very sensitive to contradictory evidence, usually rendering counter-intuitive results (which is the second challenge mentioned above). We address challenges (1) and (2) by performing fusion operations based on the CUE (Wickramarathne, et. al., 2010).

Computational complexity (i.e., challenge (3)) of DS methods exponentially increases with increasing cardinality of the FoD. As a result, in many DS-based applications, even the most common and fundamental task of conditioning can quickly become computationally prohibitive, especially in the presence of FoDs with high cardinality. To reduce computational complexity we make use of the Conditional Core Theorem (CCT) (Wickramarathne, Premaratne, and Murthi, 2010). The CCT identifies the propositions that will receive a positive mass after conditioning without any numerical computations. The advantage of such a result is that it is possible to avoid the computation of all the $2^{|\Theta|}$ propositions that otherwise would have to be computed to evaluate the conditional masses. In real application settings, the CCT may yield computational savings of 80% or more.

It is worth noting that CUE-based operations are embedded into the credibility estimation method described in Section 3 above. Then, in our example, meaning extraction could be obtained from analyzing the consensus BoE defined by $m_{\text{consensus}}(\ \{\text{white}\} \times \{\text{sedan}\}\ ) = 1.00$. In this case, the result indicates that there is total certainty that the suspect was driving a white sedan. Inference and meaning extraction in more complex scenarios can be done by following the methodology introduced in (Wickramarathne, et. al., 2011).

5 CONCLUSIONS

In this paper we introduced a general framework that allows automatic and efficient processing of both hard and soft information, leading to enhanced decision-making in multi-source environments. In this framework, soft data is converted into FOL constructs by a NLP module. This module consists of part-of-speech tagging, dependency parsing, coreference resolution, and a conversion to semantics. The logic constructs resulting from this module are quantified and mathematically modeled. In particular, DS theoretic models are generated based on methods arising from the CUE properly tuned for consistency with logic operations. The mathematical models are then used for assessing credibility. The latter is estimated based on the distance to the consensus among information sources. The credibility measure is then used for discounting BoEs in DS-based information fusion. The overall framework could be directly applied for solving problems in varied areas such as neuroscience, business analytics, military intelligence, and sociology, among others.

ACKNOWLEDGMENTS

This work is based on research supported by the US Office of Naval Research (ONR) via grants #N00014-10-1-0140 and #N00014-11-1-0493, and the US National Science Foundation (NSF) via grant #1038257.

REFERENCES

Benavoli, A., L. Chisci, A. Farina, B. Ristic, 2008. Modelling Uncertain Implication Rules in Evidence Theory. 11th International Conference on Information Fusion, pp.1-7, June 30 2008-July 3 2008

Cantrell, R. 2009. Mink: An Incremental Data-Driven Dependency Parser with Integrated Conversion to Semantics. Student Workshop at RANLP, Borovets, Bulgaria.

Genesereth, M. R., N. Nilsson, 1987. Logical Foundations of Artificial Intelligence. Morgan Kaufmann Publishers: San Mateo, CA.

Hewawasam, K.K.R., K. Premaratne, S. Mei-Ling, 2007. Rule Mining and Classification in a Situation Assessment Application: A Belief-Theoretic Approach for Handling Data

Imperfections, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol.37, no.6, pp.1446-1459, Dec. 2007

Kübler, S., M. Scheutz, E. Baucom, and R. Israel. 2010. Adding Context Information to Part Of Speech Tagging for Dialogues. Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT). Tartu, Estonia.

Kübler, S., E. Baucom. 2011. Fast Domain Adaptation for Part of Speech Tagging for Dialogues. Poceedings of the International Conference on Recent Advances in NLP (RANLP), Hissar, Bulgaria.

Kübler, S., R. Cantrell, M. Scheutz, 2011. Actions Speak Louder than Words: Evaluating Parsers in the Context of Natural Language Understanding Systems for Human-Robot Interaction. Proceedings of the International Conference on Recent Advances in NLP (RANLP), Hissar, Bulgaria.

Marcus, M., B. Santorini, and M. A. Murthi. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics 19(2), pp. 313-330.

Nivre, J., J Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. MaltParser: A Language-Independent System for Data-Driven Dependency Parsing. Natural Language Engineering 13(2), 95-135.

Pravia, M.A., R. K. Prasanth, P. O. Arambel, C. Sidner, C. Chee-Yee, 2008. Generation of a Fundamental Data Set for Hard/Soft Information Fusion. 11th International Conference on Information Fusion, pp.1-8, June 30 2008-July 3 2008

Premaratne, K., M. N. Murthi, J. Zhang, M. Scheutz, P. H. Bauer, 2009. A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data. Procedings of the International Conference on Information Fusion (ICIF'09), pp. 2122–2129. Seattle, WA 2009

Ristic, B., P. Smets, 2005. Target Identification Using Belief Functions and Implication Rules. IEEE Transactions on Aerospace and Electronic Systems, vol.41, no.3, pp.1097-1103, July 2005

Shafer, G. 1976. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press.

Wickramarathne, T. L., K. Premaratne, M. N. Murthi, and M. Scheutz, 2010. A Dempster-Shafer theoretic evidence updating strategy for non-identical frames of discernment, In Proc. Workshop on the Theory of Belief Functions (WTBF), Brest, France, Apr. 2010.

Wickramarathne, T. L., K Premaratne, M. N. Murthi, 2010. Focal elements generated by the Dempster-Shafer theoretic conditionals: a complete characterization. International Conference on Information Fusion (FUSION), Edinburgh, UK, pp. 1-8, July 2010

Wickramarathne, T.L., K. Premaratne, M. N. Murthi, M. Scheutz, S. Kübler, M. Pravia, 2011. Belief Theoretic Methods for Soft and Hard Data Fusion. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.2388-2391, 22-27 May 2011

Wickramarathne, T. L., K. Premaratne, and M. N. Murthi. 2012. Consensus-Based Credibility Estimation of Soft Evidence for Robust Data Fusion. 2nd International Conference on Belief Functions, Compiegne, France, 2012.

Zhekova, D. and S. Kübler. 2010. UBIU: A Language-Independent System for Coreference Resolution. Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval), Uppsala, Sweden.

Zhekova, D. and S. Kübler. 2011. UBIU: A Robust System for Resolving Unrestricted Coreference. Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL), Shared Task, Portland, OR.