

## REFERENCES

- [1] Amit Singhal, "Modern Information Retrieval: A Brief Overview," *BULLETIN OF THE IEEE COMPUTER SOCIETY TECHNICAL COMMITTEE ON DATA ENGINEERING*, 2001.
- [2] J Cowie and Y Wilks, "Information Extraction," , 1996.
- [3] Daya C. Wimalasuriya and Dejing Dou, "Ontology-based information extraction: An introduction and a survey of current approaches," *Journal of Information Science* 2010, vol. 36, no. 3, pp. 306-323, June 2010.
- [4] R Studer, V.R Benjamins, and D Fensel, "Knowledge Engineering: Principles and Methods," in *Data and Knowledge Engineering* 25, 1998, pp. 161-197.
- [5] Daya C. Wimalasuriya and Dejing Dou, "Components for Information Extraction: Ontology-Based Information Extractors and Generic Platforms," in *CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 9-18.
- [6] M. Gruninger and M. Uschold, "Ontologies: Principles, Methods and Applications," *Knowledge Engineering Review*, vol. 11, no. 2, pp. 93-136, June 1996.
- [7] N. Guarino, D. Oberle, and S. Staab, "What Is an Ontology?," in *Handbook on Ontologies, International Handbooks on Information Systems.*: Springer-Verlag Berlin Heidelberg, 2009.
- [8] Noy, Natalya F and McGuinness, Deborah L.. (2001, March) *Ontology Development 101: A Guide to Creating Your First Ontology*. [Online]. Available: [http://protege.stanford.edu/publications/ontology\\_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html)
- [9] Carlos Vicient Monllaó, "Ontology-based Information Extraction," Interuniversity Master in Artificial Intelligence (UPC-URV-UB), LSI Department, Technical University of Catalonia, Master of Science Thesis 2011.
- [10] Tim Berners-Lee, James Hendler, and Ora Lassila. (2001, May) The Semantic Web. Document.
- [11] Daya C. Wimalasuriya and Dejing Dou, "Using Multiple Ontologies in Information Extraction," in *18th ACM conference on Information and knowledge management*, 2009, pp. 235-244.

- [12] W3C. (2004, February) *Web Service Glossary, W3C Working Group Note 11*. [Online]. Available: <http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice>
- [13] W3C. (2004, February) *Web Service Architecture, W3C Working Group Note 11*. [Online]. Available: <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/#gengag>
- [14] Einar Landre and Harald Wesenberg, "REST versus SOAP as Architectural Style for Web Services," in *ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages and Applications*, Montréal, Québec, Canada, 2007.
- [15] Hao He. (2003, September) What Is Service-Oriented Architecture. Document.
- [16] Xinyang Feng, Jianjing Shen, and Ying Fan, "REST : An Alternative to RPC for Web Services," in *ICFIN 2009, First International Conference on Future Information Networks.*, Beijing , 2009, pp. 7-10.
- [17] W3C. (2000, May) *Simple Object Access Protocol (SOAP) 1.1*, *W3C Note 08 May 2000*. [Online]. Available: <http://www.w3.org/TR/2000/NOTE-SOAP-20000508/>
- [18] P.A. Castillo, J.L. Bernier, M.G. Arenas, J.J. Merelo, and P. Garcia-Sanchez, "SOAP vs REST: Comparing a master-slave CA implementation," in *First International Workshop of Distributed Evolutionary Computation in Informatics*, New Orleans, 2011.
- [19] Roy Thomas Fielding, "Architectural Styles and the Design of Network-based Software Architectures," University of California, Irvine, Doctoral dissertation 2000.
- [20] M. zur Muehlen, J.V. Nickerson, and K.D. Swenson, "Developing Web Services Choreography Standards – The Case of REST vs. SOAP," *Decision Support Systems - Special issue: Web services and process management*, vol. 40, no. 1, pp. 9-29, July 2005.
- [21] G. Mulligan and D. Gracanin, "A Comparison of SOAP and REST Implementation of a Service Based Interaction Independence Middleware Framework," in *Winter Simulation Conference*, Austin, TX , 2009, pp. 1423-1432.
- [22] Hamish Cunningham, Valenti Tablan, Angus Roberts, and Kalina Bontcheva, "Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics," *PLOS Computational Biology*, vol. 9, no. 2, February 2013.
- [23] René Witte, Ninus Khamis, and Juergen Rilling, "Flexible Ontology Population from Text: The OwlExporter," in *Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010.

- [24] Paul Buitelaar, Philipp Cimiano, Stefania Racioppa, and Melanie Siegel, "Ontology-based Information Extraction with SOBA," in *LREC*, Genoa, Italy, 2006.
- [25] (2015) *GATE General Architecture for Text Engineering, Chapter 8, JAPE:Regular Expressions over Annotations.* [Online]. Available: <https://gate.ac.uk/sale/tao/splitch8.html>
- [26] Hamish Cunningham, Kevin Humphreys, Robert Gaizauskas, and Yorick Wilks, "GATE - A General Architecture for Text Engineering," in *Proceedings of the Fifth Conference on Applied Natural Language Processing: Descriptions of System Demonstrations and Videos*, Washington, DC, USA, 1997, pp. 29--30.
- [27] Ian H Witten et al., "Weka: Practical Machine Learning Tools and Techniques with Java Implementations," , 1999.
- [28] A. K. McCallum. (2002) *MALLET: A Machine Learning for Language Toolkit.* [Online]. Available: <http://mallet.cs.umass.edu>
- [29] (2015) *Java API for RESTful Web Services.* [Online]. Available: [http://en.wikipedia.org/wiki/Java\\_API\\_for\\_RESTful\\_Web\\_Services](http://en.wikipedia.org/wiki/Java_API_for_RESTful_Web_Services)
- [30] Java.net. (2014) *Java API for RESTful Services (JAX-RS).* [Online]. Available: <https://jax-rs-spec.java.net/>
- [31] (2010, April) *Apache Wink 1.4.0 user guide.* [Online]. Available: [https://wink.apache.org/1.4.0/Apache\\_Wink\\_User\\_Guide.pdf](https://wink.apache.org/1.4.0/Apache_Wink_User_Guide.pdf)
- [32] Vishnu Vettrival. (2010, April) *RESTful Web services with Apache Wink, Part 3: Apache Wink and the REST - A comparison of Apache Wink and other open source JAX-RS implementations.* [Online]. Available: <http://www.ibm.com/developerworks/library/wa-apachewink3/>
- [33] (2015) *Jersey: RESTful web services in java.* [Online]. Available: <https://jersey.java.net/>
- [34] *RestEasy:A jboss project.* [Online]. Available: <http://resteasy.jboss.org/>
- [35] *Restlet Framework.* [Online]. Available: <http://restlet.com/products/restlet-framework/>
- [36] Kalina Bontcheva, Valentin Tablan, Diana Maynard, and Hamish Cunningham, "Evolving GATE to Meet New Challenges in Language Engineering," *Natural Language Engineering*, vol. 10, no. 3/4, pp. 349-373, 2004.

## APPENDIX A: Sample Metadata Files – Information Extractor

### Extract Currency Information

```
<?xml version="1.0"?>
<InformationExtractor
xmlns="http://aimlab.cs.uoregon.edu/obie/StandardXMLSchema/1.0/">
<Language>EN</Language>
<IETechnique>http://aimlab.cs.uoregon.edu/obie/IETechniques/LinguisticExtractionR
ules.xml</IETechnique>
<Concept>Currency</Concept>
<ConceptType>Class</ConceptType>
<IdentifierName>hasName</IdentifierName>
<Feature>rule:input:Lookup</Feature>
<Feature>rule:
    ({Lookup.minorType == post_amount})
    : curr -->
    {
        gate.AnnotationSet curSet = (gate.AnnotationSet)bindings.get("curr");
        FeatureMap features = Factory.newFeatureMap();
        outputAS.add(curSet.firstNode(), curSet.lastNode(),"CurrencyX", features );
    }
</Feature>
</InformationExtractor>
```



University of Moratuwa, Sri Lanka.  
Electronic Theses & Dissertations  
[www.lib.mrt.ac.lk](http://www.lib.mrt.ac.lk)

Extract the names of Capitals mentioned in the corpus

```
<?xml version="1.0"?>

<InformationExtractor
xmlns="http://aimlab.cs.uoregon.edu/obie/StandardXMLSchema/1.0/">

<Language>EN</Language>

<IETechnique>http://aimlab.cs.uoregon.edu/obie/IETechniques/LinguisticExtractionR
ules.xml</IETechnique>

<Concept>Capital</Concept>

<ConceptType>Class</ConceptType>

<IdentifierName>hasName</IdentifierName>

<Feature>rule-input:Lookup</Feature>

<Feature>rule:
({Lookup.minorType == city}).
: loc -->
{
    gate.AnnotationSet curSet = (gate.AnnotationSet)bindings.get("loc");
    FeatureMap features = Factory.newFeatureMap();
    outputAS.add(curSet.firstNode(), curSet.lastNode(), "CapitalX", features );
}
</Feature>

</InformationExtractor>
```



University of Moratuwa, Sri Lanka.  
Electronic Theses & Dissertations  
[www.lib.mrt.ac.lk](http://www.lib.mrt.ac.lk)

## APPENDIX B: Sample Metadata File – Two-Phase Classifier

To identify the Perpetrators given in the corpus

```
<?xml version="1.0" encoding="UTF-8"?>
<InformationExtractor
xmlns="http://aimlab.cs.uoregon.edu/obie/StandardXMLSchema/1.0/">
<Language>EN</Language>
<IETechnique>http://aimlab.cs.uoregon.edu/obie/IETechniques/TwoStepClassification.xml</IETechnique>
<Concept>Perpetrator</Concept>
<ConceptType>Class</ConceptType>
<IdentifierName>hasName</IdentifierName>
<Feature>word:attack</Feature>
<Feature>word:kidnap</Feature>
<Feature>word:kill</Feature>
<Feature>word:threaten</Feature>
<Feature>word:burn</Feature>
<Feature>word:civilian</Feature>
<Feature>word:peasant</Feature>
<Feature>word:leader</Feature>
<Feature>word:politician</Feature>
<Feature>word:officer</Feature>
<Feature>word:military</Feature>
<Feature>word:innocent</Feature>
<Feature>word:terrorist</Feature>
<Feature>word:soldier</Feature>
<Feature>word:guerrilla</Feature>
<Feature>word:armedforces</Feature>
<Feature>word:armed</Feature>
<Feature>word:individual</Feature>
<Feature>word:suspected</Feature>
<Feature>word:commando</Feature>
```



University of Moratuwa, Sri Lanka.  
Electronic Theses & Dissertations  
www.lib.mrt.ac.lk

<Feature>word:officers</Feature>  
<Feature>word:army</Feature>  
<Feature>word:murderer</Feature>  
<Feature>word:leftist</Feature>  
<Feature>word:rightwing</Feature>  
<Feature>word:kidnapper</Feature>  
<Feature>word:navy</Feature>  
<Feature>word:airforce</Feature>  
<Feature>word:revolutionary</Feature>  
<Feature>word:movement</Feature>  
<Feature>word:spokesman</Feature>  
<Feature>word:spokeswoman</Feature>  
<Feature>word:clandestine</Feature>  
<Feature>word:arson</Feature>  
<Feature>word:shoot</Feature>  
<Feature>word:dynamite</Feature>  
<Feature>word:tnt</Feature>  
<Feature>word:explosive</Feature>  
<Feature>word:grenade</Feature>  
<Feature>word:bullet</Feature>  
<Feature>word:axe</Feature>  
<Feature>word:fuse</Feature>  
<Feature>synset:attack</Feature>  
<Feature>synset:kidnap</Feature>  
<Feature>synset:kill</Feature>  
<Feature>synset:threaten</Feature>  
<Feature>synset:burn</Feature>  
<Feature>synset:civilian</Feature>  
<Feature>synset:peasant</Feature>  
<Feature>synset:leader</Feature>  
<Feature>synset:politician</Feature>  
<Feature>synset:officer</Feature>



<Feature>synset:military</Feature>  
<Feature>synset:innocent</Feature>  
<Feature>synset:shoot</Feature>  
<Feature>gazetteer:firstname</Feature>  
<Feature>gazetteer:lastname</Feature>  
<Feature>gazetteer:military\_rank</Feature>  
<Feature>gazetteer:job\_title</Feature>  
<Feature>gazetteer:terrorist\_organization</Feature>  
<Feature>gazetteer:instrument\_type</Feature>  
</InformationExtractor>



University of Moratuwa, Sri Lanka.  
Electronic Theses & Dissertations  
[www.lib.mrt.ac.lk](http://www.lib.mrt.ac.lk)