Research Interests

Louise Guthrie

_____

Overview

Most of my research has been in using mathematical techniques to improve how computers process or understand natural language. Some of the topics that I've worked on over the years are:

- Machine Translation
- Automatically extracting important information from text
- Information Retrieval
- Making use of the electronic forms of dictionaries to help computers "understand text"
- Classifying texts by topic or genre or authorship
- Finding anomalous information in text

Below is a list of some of my publications and research contracts that I've worked on

Books

Wilks, Y., Slator, B. and Guthrie, L. (1996)
Electric Words: Dictionaries, Computers, and Meanings MIT Press, 288 pp.

Selected Papers

Guthrie, D., Guthrie, L., and Wilks, Y. (2009)
"What is a "full statistical model" of a language and are there short cuts to it?".
In *After Half a Century of Slavonic Natural Language Processing,* Eds D. Hilavackova, A. Horak, K. Osolsobe, P. Rychly, pp. 45-56

Jabbari, S., Guthrie, L. (2009)
"A Probabilistic Model of Context Applied to the Lexical Substitution Task".
In Proceedings of the 12th International Conference on Text, Speech and Dialogue conference
(TSD 09), Pilzen, Czech Republic pp 427-436

Liu, W.., Guthrie, L. (2009)
"Chinese Pinyin Text Conversion on Segmented Text".
In Proceedings of the 12th International Conference on Text, Speech and Dialogue conference
(TSD 09), Pilzen, Czech Republic pp 427-436

Jabbari, S. & Allison, B., and Guthrie, L. (2008)
"An Empirical Bayesian Method for Detecting Out Of Context Words".
In Proceedings TSD'08, Brno, Czech Republic

Allison, B., & Guthrie, L. (2008)
"Authorship Attribution of E-Mail: Comparing Classifiers Over a New Corpus for Evaluation"
In Proceedings LREC'08, Marrakech, Morocco

Jabbari, S., Allison, B., and Guthrie, L., (2008)
"Using a Probabilistic Model of Context to Detect Word Obfuscation".
In Proceedings LREC'08, Marrakech, Morocco

Liu, W., Allison, B., and Guthrie, L., (2008)
"Professor or screaming beast? Detecting Anomalous Words in Chinese"
In Proceedings LREC'08, Marrakech, Morocco

Guthrie, D., Guthrie, L., Allison, B., and Wilks, Y., (2007)
"Unsupervised Anomaly Detection"
In Proceedings IJCAI'07, Hyderbad, India, pp. 1624-1628

Jabbari, S., Allison, B., Guthrie, D. & Guthrie, L. (2006)
"Towards the Orwellian Nightmare: Separation of Business and Personal Emails"
In Proceedings of 21st international Conference for Computational Linguistics and
44th Annual meeting of Association for Computational Linguistics, pp. 407-411
Sydney Australia, July 2006

Allison, B., Guthrie, D.& Guthrie, L. (2006)
"Another Look at the Data Sparsity Problem"
In Proceedings of the Ninth International Conference on Text, Speech and Dialogue
Brno, Czech Republic September, 2006

David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, Yorick Wilks (2006)
"A Closer Look at Skip-gram Modelling"
 In Proceedings of the Fifth international Conference on Language Resources and
Evaluation, pp. 1222-1225, Genoa, Italy May 200Louise Guthrie, Wei Liu, Yunqing Xia, (2005)

"Yunqing Xia, Wei Liu, Louise Guthrie, (2005]
"Email Categorization with Tournament Methods."
In Proceedings of the 10th International Conference on Applications of Natural Language to
Information SystemsAlicante, Spain, June 2005, , pp 150-160

Yunqing Xia, Angelo Dalli, Yorick Wilks, Louise Guthrie (2005)
"FASiL Adaptive Email Categorisation System",
 In Lecture Notes in Computer Science 3406,
 Proceedings of the 6th International Conference on Computational Linguistics and
Intelligent Text Processing, Mexico City, Mexico, Feb. 2005, , pp. 723-734.

Yunqing Xia, Wei Liu,Louise Guthrie,Kam-Fai Wong, (2005)
"Improving Quality of Email Categorization with Tournament Methods"
In Proceedings of the International Conference on Information and Knowledge Engineering,
Las Vegas, Nevada, June2005. pp215-221

Louise Guthrie, Roberto Basili, Fabio Zanzotto, Kalina Bontcheva, Hamish Cunningham,
David Guthrie, Jia Cui, Marco Cammisa, Jerry Cheng-Chieh Liu, Cassia Farria Martin,
Krystian Haralambiev, Martin Holub, Klaus Macherey, Fredreick Jelinek (2004);
"Large Scale experiments for Semantic Labeling of Noun Phrases in Raw Text"
In the Proceedings of the 4th International Conference on Language Resources and
Evaluation LREC 2004, Lisbon Portugal, May 2004, pp. 26-27

Peters, W., Guthrie, L., Wilks, Y. (2002),
"Computational Psycholinguistics: The Case of Regular Polysemy",
In: Petrova K., Eftimova A., Shopov R. (Eds.), Litora Psycholinguistica,
SEMA RSH, Sofia 2002, pp. 328-338.

Louise Guthrie, Wei Liu, Yunqing Xia, (2005)
Text Classification with Tournament Methods"
In Proceedings of the Eighth International Conference on Text, Speech and Dialogue
Carlsbad, Czech Republic September 2005

Guthrie, L., Leistensnider, J., and Guthrie, J. (1999)
"Document Classification and Routing: a Probabilistic Approach";
Natural Language Information Retrieval, T. Strzalkowski (ed.),   Kluwer Academic Publishers,
Dordrecht, The Netherlands, pp.289-306.

Guthrie, L., Pustejovsky, J., Slator, B. and Wilks, Y. (1996)
"The Role of Lexicons in Natural Language Processing".
Communications of the ACM, Special Issue on Natural Language Processing, January 1996,
Vol. 39 No. 1, pp. 63-72.

Guthrie, L. and Strzalkowski, T., Lin, F. and Wang, J.(1996)
"Integration of Document Detection and Information Extraction";
Invited Paper to describe U.S. Government funded research.
Proceedings of the Tipster Progam Phase II: Advances in Text Processing, Vienna, Virginia;
Morgan Kaufmann, pp. 195-200.

Guthrie, L. and Leistensnider, J., (1996)
"A Simple Probabilistic Approach to Classification and Routing" ;
Invited Paper to describe U.S. Government funded research.
Proceedings of the Tipster Progam Phase II: Advances in Text Processing, Vienna, Virginia;
Morgan Kaufmann, pp. 167-178.

Cowie, J., Guthrie, J. and Guthrie, L. (1992)
"Lexical Disambiguation using Simulated Annealing";
Proceedings of the 16th International Conference on Computational Linguistics (COLING-92)
Nantes, France, July, pp. 359-365. [A description of preliminary results appears in the
Proceedings of the Fifth DARPA Speech and Natural Language Workshop, February 1992, pp.
238-242.]

Guthrie, J., Guthrie, L., Wilks, Y., and Aidinejad, H. (1991).
 "Subject-Dependent Co-Occurrence and Word Sense Disambiguation".
Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics,
Berkeley, CA. June. pp. 146-152.

Funded Research

Distinguishing the expert and non-expert in text - GTW partners £145 - July 2011-2012 (co-PI)

REVEAL II   The identification of anomalous segments in text on a large scale
  (2008-2010), MOD funded University of Sheffield, PI
Budget amount about £300,000.

CLUE : Contextual Learning for detecting Unexpexted Events (2006-2008)
MOD funded; University of Sheffield, PI
Budget amount: £360,854

PEEC : Partitioning the Enron Email Corpus (A proposal to annotate personal and business emails) (2006-2007)
MOD funded , University of Sheffield, PI
Budget amount: £119, 879

ABRAXAS : Automating Ontology Learning for the Semantic Web (2004-2007)
EPSRC funded, University of Sheffield, co-PI
Budget amount: £164,614, (co-PI) (about
COMPANIONS : Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet (2006-2010)
EU funded, Co-ordinator: University of Sheffield, (co-PI)
University of Sheffield Budget amount - £1,599,965 (total funding amount €10,178,219) (about $3,000,000 for Sheffield, and about $15, 000,000 overall)

REVEAL   The identification of anamalous segments in text on a large scale
  (2005-2007)
MOD funded; University of Sheffield (co- PI)
Budget amount: £149,894

Genre Identification Development and Expansion (2000 - 2001)
Department of Defense; New Contract at University of Texas at El Paso (PI)
Budget amount: $75K

Content Representation Assessment (2000 - 2001)
Department of Defense; New Contract at University of Texas at El Paso (PI)
Budget amount: $150K

Consulting on LDRD Project (1997)
Sandia National Laboratories; Computing Research Laboratory , New
Mexico State University (PI)
Budget amount: $15K)

Text Classification (1995-1996)
ARPA: Lockheed Martin Management and Data Systems, (PI)
Budget amount: $50K

Lockheed Martin TIPSTER Phase II (1994-1996)
ARPA: Lockheed Martin Management and Data Systems,   (PI)
Budget amount: $2million

Cibola/Oleada: Translator Support Systems (1994-1998)
DOD; Computing Research Laboratory, New Mexico State University (consultant)
Budget amount: $900K

Tipster II: Cervantes (1994-1996)
DARPA; Computing Research Laboratory, New Mexico State University (consultant)
Budget amount: $400K)

CRL Translation Support (1994-1995)
IRS: Computing Research Laboratory, New Mexico State University (PI)
Budget amount: $100K

Improvement of Word Sense Info. in the English Word Dict. II (1993-1994)
EDR; Computing Research Laboratory, New Mexico State University (PI)
Budget amount: $75K

A Lexical Knowledge Base from a Machine Readable Dictionary (1991-1995)
NSF; CRL contract   134 (co-PI)

Budget amount: $900K

DARPA MT Phase 2 Knowledged-Based Machine Translation Project (1992-1994)
DARPA; Computing Research Laboratory, New Mexico State University (consultant-automatic construction of lexicons)
Budget amount: $600K

EDR Dictionary Evaluation (1992-1994)
EDR; Computing Research Laboratory, New Mexico State University (co-PI)
Budget amount: $100K.

TIPSTER Extraction Project (1991-1994)
DARPA; Computing Research Laboratory, New Mexico State University (consultant - all aspects)
Budget amount: $800K

Consortium for Lexical Research (1991-1993)
DARPA: Computing Research Laboratory, New Mexico State University (co-PI)
Budget amount: $100K

Knowledge-based Machine Translation Research Project: Site 2 (1991-1992)
DARPA; Computing Research Laboratory, New Mexico State University (consultant-automatic construction of lexicons)
Budget amount: $400K

General Operating System Consultant (1988-1990)
US West Adv. Technologies: Computing Research Laboratory, New Mexico State University (co-PI)
 Budget amount: $200K