Michael O'Neill
Richard F. E. Sutcliffe
Conor Ryan
Malachy Eaton
Niall J. L. Griffith (Eds.)

LNAI 2464

Artificial Intelligence and Cognitive Science

13th Irish Conference, AICS 2002
Limerick, Ireland, September 2002
Proceedings

Springer

Michael O'Neill Richard F. E. Sutcliffe
Conor Ryan Malachy Eaton
Niall J. L. Griffith (Eds.)

Artificial Intelligence and Cognitive Science

13th Irish Conference, AICS 2002
Limerick, Ireland, September 12-13, 2002
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Michael O'Neill
Richard F. E. Sutcliffe
Conor Ryan
Malachy Eaton
Niall J. L. Griffith
University of Limerick
Department of Computer Science and Information Systems
Ireland
{michael.oneill, richard.sutcliffe, conor.ryan, malachy.eaton, niall.griffith}@ul.ie

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Artificial intelligence and cognitive science : 13th Irish conference ; proceedings / AICS 2002, Limerick, Ireland, September 12 - 13, 2002. Michael O'Neill ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2002
(Lecture notes in computer science ; 2464 : Lecture notes in artificial intelligence)
ISBN 3-540-44184-0

CR Subject Classification (1998): I.2, F.1

ISSN 0302-9743
ISBN 3-540-44184-0 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York,
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna e.K.
Printed on acid-free paper SPIN: 10871241 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 2464

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Preface

The Artificial Intelligence and Cognitive Science Conference has taken place annually since 1988. It provides a forum for the exchange of ideas and the presentation of results relating to work conducted both in Ireland and worldwide.

The conference spans a large number of fields including case-based reasoning, cognitive modeling, constraint processing, data mining, evolutionary computation, intelligent agents, intelligent information retrieval, knowledge representation and reasoning, learning, natural language processing, neural networks, perception and planning, robotics, and scheduling.

AICS 2002 was the thirteenth conference in the series and took place at University of Limerick on 12–13 September. In addition to the 16 regular papers and 17 concise papers accepted for presentation, we were delighted to welcome two prestigious keynote speakers both fitting in with the conference theme 'Towards Bio-inspired Computing'. They were David Goldberg of University of Illinois at Urbana-Champaign, and Dario Floreano of the Swiss Federal Institute of Technology.

I would like to take this opportunity to thank our sponsors QAD Ireland and University of Limerick, as well as all those who were involved in the organisation of the conference including the Co-chairs, Programme Committee members, and Conference Administrators.

June 2002

Richard F. E. Sutcliffe

Organisation

AICS 2002 was organized by the Department of Computer Science and Information Systems, University of Limerick, in association with the Artificial Intelligence Association of Ireland.

Organising Committee

General Chair:	Richard F.E. Sutcliffe
Publication Chair:	Michael O'Neill
Program Co-chair:	Conor Ryan
Local Arrangements Chair:	Malachy Eaton
Program Co-chair:	Niall Griffth
Conference Administrators:	Maeve Gleeson, Bianca Pluk

Programme Committee

Liam Bannon, University of Limerick
David Bell, University of Ulster at Jordanstown
Anthony Brabazon, University College Dublin
Derek Bridge, University College Cork
Jim Buckley, University of Limerick
Ruth Byrne, The University of Dublin, Trinity College
Arthur Cater, University College Dublin
J.J. Collins, University of Limerick
Ernesto Costa, University of Coimbra
Fintan Costello, Dublin City University
Roddy Cowie, Queen's University Belfast
Norman Creaney, University of Ulster at Coleraine
Fred Cummins, University College Dublin
Pádraig Cunningham, The University of Dublin, Trinity College
John Dunnion, University College Dublin
Jim Eales, University of Limerick
Malachy Eaton, University of Limerick
José Luis Fernández-Villacañas, Universidad Carlos III de Madrid
Dario Floreano, Swiss Federal Institute of Technology
Antony Galton, University of Exeter
Paul Gibson, National University of Ireland, Maynooth
David Goldberg, University of Illinois at Urbana-Champaign
Josephine Griffith, National University of Ireland, Galway
Niall Griffith, University of Limerick
Jane Grimson, The University of Dublin, Trinity College

Patrick Healy, University of Limerick
Michael Herdy, INPRO
Gareth Jones, University of Exeter
Mark Keane, University College Dublin
Maarten Keijzer, Vrije University, Amsterdam
Nick Kushmerick, University College Dublin
Michael Luck, University of Southampton
Annette McElligott, University of Limerick
Paul Mc Kevitt, University of Ulster at Magee
David McSherry, University of Ulster at Coleraine
Mike McTear, University of Ulster at Jordanstown
Juan Julián Merelo, University of Granada
Liam Murray, University of Limerick
Fionn Murtagh, Queen's University Belfast
Ajit Narayanan, University of Exeter
Diarmuid O'Donoghue, National University of Ireland, Maynooth
Greg O'Hare, University College Dublin
Donncha Ó Maidín, University of Limerick
Michael O'Neill, University of Limerick
Colm O'Riordan, National University of Ireland, Galway
Riccardo Poli, University of Essex
Ronan Reilly, National University of Ireland, Maynooth
Conor Ryan, University of Limerick
Reinhard Schaeler, University of Limerick
Jonathan Shapiro, University of Manchester
Noel Sharkey, University of Sheffield
Barry Smyth, University College Dublin
Humphrey Sorensen, University College Cork
Richard F. E. Sutcliffe, University of Limerick
Adrian Trenaman, Iona Technologies
Josef van Genabith, Dublin City University
Carl Vogel, The University of Dublin, Trinity College
Andy Way, Dublin City University
Adam Winstanley, National University of Ireland, Maynooth

Sponsors

QAD Ireland
University of Limerick

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 2317: M. Hegarty, B. Meyer, N. Hari Narayanan (Eds.), Diagrammatic Representation and Inference. Proceedings, 2002. XIV, 362 pages. 2002.

Vol. 2321: P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Advances in Learning Classifier Systems. Proceedings, 2002. VIII, 231 pages. 2002.

Vol. 2322: V. Mařík, O. Stěpánková, H. Krautwurmová, M. Luck (Eds.), Multi-Agent Systems and Applications II. Proceedings, 2001. XII, 377 pages. 2002.

Vol. 2432: R. Bergmann, Experience Management. XXI, 393 pages. 2002.

Vol. 2333: J.-J.Ch. Meyer, M. Tambe (Eds.), Intelligent Agents VIII. Revised Papers, 2001. XI, 461 pages. 2001.

Vol. 2336: M.-S. Chen, P.S. Yu, B. Liu (Eds.), Advances in Knowledge Discovery and Data Mining. Proceedings, 2002. XIII, 568 pages. 2002.

Vol. 2338: R. Cohen, B. Spencer (Eds.), Advances in Artificial Intelligence. Proceedings, 2002. XII, 373 pages. 2002.

Vol. 2356: R. Kohavi, B.M. Masand, M. Spiliopoulou, J. Srivastava (Eds.), WEBKDD 2002 – Mining Web Log Data Across All Customers Touch Points. Proceedings, 2002. XI, 167 pages. 2002.

Vol. 2358: T. Hendtlass, M. Ali (Eds.), Developments in Applied Artificial Intelligence. Proceedings, 2002 XIII, 833 pages. 2002.

Vol. 2366: M.-S. Hacid, Z.W. Raś, D.A. Zighed, Y. Kodratoff (Eds.), Foundations of Intelligent Systems. Proceedings, 2002. XII, 614 pages. 2002.

Vol. 2371: S. Koenig, R.C. Holte (Eds.), Abstraction, Reformulation, and Approximation. Proceedings, 2002. XI, 349 pages. 2002.

Vol. 2375: J. Kivinen, R.H. Sloan (Eds.), Computational Learning Theory. Proceedings, 2002. XI, 397 pages. 2002.

Vol. 2377: A. Birk, S. Coradeschi, T. Satoshi (Eds.), RoboCup 2001: Robot Soccer World Cup V. XIX, 763 pages. 2002.

Vol. 2381: U. Egly, C.G. Fermüller (Eds.), Automated Reasoning with Analytic Tableaux and Related Methods. Proceedings, 2002. X, 341 pages. 2002 .

Vol. 2385: J. Calmet, B. Benhamou, O. Caprotti, L. Henocque, V. Sorge (Eds.), Artificial Intelligence, Automated Reasoning, and Symbolic Computation. Proceedings, 2002. XI, 343 pages. 2002.

Vol. 2389: E. Ranchhod, N.J. Mamede (Eds.), Advances in Natural Language Processing. Proceedings, 2002. XII, 275 pages. 2002.

Vol. 2392: A. Voronkov (Ed.), Automated Deduction – CADE-18. Proceedings, 2002. XII, 534 pages. 2002.

Vol. 2393: U. Priss, D. Corbett, G. Angelova (Eds.), Conceptual Structures: Integration and Interfaces. Proceedings, 2002. XI, 397 pages. 2002.

Vol. 2394: P. Perner (Ed.), Advances in Data Mining. VII, 109 pages. 2002.

Vol. 2403: Mark d'Inverno, M. Luck, M. Fisher, C. Preist (Eds.), Foundations and Applications of Multi-Agent Systems. Proceedings, 1996-2000. X, 261 pages. 2002.

Vol. 2407: A.C. Kakas, F. Sadri (Eds.), Computational Logic: Logic Programming and Beyond. Part I. XII, 678 pages. 2002.

Vol. 2408: A.C. Kakas, F. Sadri (Eds.), Computational Logic: Logic Programming and Beyond. Part II. XII, 628 pages. 2002.

Vol. 2413: K. Kuwabara, J. Lee (Eds.), Intelligent Agents and Multi-Agent Systems. Proceedings, 2002. X, 221 pages. 2002.

Vol. 2416: S. Craw, A. Preece (Eds.), Advances in Case-Based Reasoning. Proceedings, 2002. XII, 656 pages. 2002.

Vol. 2417: M. Ishizuka, A. Sattar (Eds.), PRICAI 2002: Trends in Artificial Intelligence. Proceedings, 2002. XX, 623 pages. 2002.

Vol. 2430: T. Elomaa, H. Mannila, H. Toivonen (Eds.), Machine Learning: ECML 2002. Proceedings, 2002. XIII, 532 pages. 2002.

Vol. 2431: T. Elomaa, H. Mannila, H. Toivonen (Eds.), Principles of Data Mining and Knowledge Discovery. Proceedings, 2002. XIV, 514 pages. 2002.

Vol. 2443: D. Scott (Ed.), Artificial Intelligence: Methodology, Systems, and Applications. Proceedings, 2002. X, 279 pages. 2002.

Vol. 2445: C. Anagnostopoulou, M. Ferrand, A. Smaill (Eds.), Music and Artificial Intelligence. Proceedings, 2002. VIII, 207 pages. 2002.

Vol. 2446: M. Klusch, S. Ossowski, O. Shehory (Eds.), Cooperative Information Agents VI. Proceedings, 2002. XI, 321 pages. 2002.

Vol. 2447: D.J. Hand, N.M. Adams, R.J. Bolton (Eds.), Pattern Detection and Discovery. Proceedings, 2002. XII, 227 pages. 2002.

Vol. 2448: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. Proceedings, 2002. XII, 481 pages. 2002.

Vol. 2464: M. O'Neill, R.F.E. Sutcliffe, C. Ryan, M. Eaton, N. Griffith (Eds.), Artificial Intelligence and Cognitive Science. Proceedings, 2002. XI, 247 pages. 2002.

Vol. 2479: M. Jarke, J. Koehler, G. Lakemeyer (Eds.), KI 2002: Advances in Artificial Intelligence. Proceedings, 2002. XIII, 327 pages.

Lecture Notes in Computer Science

Vol. 2422: H. Kirchner, Ch. Ringeissen (Eds.), Algebraic Methodology and Software Technology. Proceedings, 2002. XI, 503 pages. 2002.

Vol. 2423: D. Lopresti, J. Hu, R. Kashi (Eds.), Document Analysis Systems V. Proceedings, 2002. XIII, 570 pages. 2002.

Vol. 2425: Z. Bellahsène, D. Patel, C. Rolland (Eds.), Object-Oriented Information Systems. Proceedings, 2002. XIII, 550 pages. 2002.

Vol. 2426: J.-M. Bruel, Z. Bellahsène (Eds.), Advances in Object-Oriented Information Systems.Procedings, 2002. IX, 314 pages. 2002.

Vol. 2430: T. Elomaa, H. Mannila, H. Toivonen (Eds.), Machine Learning: ECML 2002. Proceedings, 2002. XIII, 532 pages. 2002. (Subseries LNAI).

Vol. 2431: T. Elomaa, H. Mannila, H. Toivonen (Eds.), Principles of Data Mining and Knowledge Discovery. Proceedings, 2002. XIV, 514 pages. 2002. (Subseries LNAI).

Vol. 2432: R. Bergmann, Experience Management. XXI, 393 pages. 2002. (Subseries LNAI).

Vol. 2434: S. Anderson, S. Bologna, M. Felici (Eds.), Computer Safety, Reliability and Security. Proceedings, 2002. XX, 347 pages. 2002.

Vol. 2435: Y. Manolopoulos, P. Návrat (Eds.), Advances in Databases and Information Systems. Proceedings, 2002. XIII, 415 pages. 2002.

Vol. 2436: J. Fong, C.T. Cheung, H.V. Leong, Q. Li (Eds.), Advances in Web-Based Learning. Proceedings, 2002. XIII, 434 pages. 2002.

Vol. 2438: M. Glesner, P. Zipf, M. Renovell (Eds.), Field-Programmable Logic and Applications. Proceedings, 2002. XXII, 1187 pages. 2002.

Vol. 2439: J.J. Merelo Guervós, P. Adamidis, H.-G. Beyer, J.-L. Fernández-Villacañas, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature – PPSN VII. Proceedings, 2002. XXII, 947 pages. 2002.

Vol. 2440: J.M. Haake, J.A. Pino (Eds.), Groupware: Design, Implementation and Use. Proceedings, 2002. XII, 285 pages. 2002.

Vol. 2442: M. Yung (Ed.), Advances in Cryptology – CRYPTO 2002. Proceedings, 2002. XIV, 627 pages. 2002.

Vol. 2443: D. Scott (Ed.), Artificial Intelligence: Methodology, Systems, and Applications. Proceedings, 2002. X, 279 pages. 2002. (Subseries LNAI).

Vol. 2444: A. Buchmann, F. Casati, L. Fiege, M.-C. Hsu, M.-C. Shan (Eds.), Technologies for E-Services. Proceedings, 2002. X, 171 pages. 2002.

Vol. 2445: C. Anagnostopoulou, M. Ferrand, A. Smaill (Eds.), Music and Artificial Intelligence. Proceedings, 2002. VIII, 207 pages. 2002. (Subseries LNAI).

Vol. 2446: M. Klusch, S. Ossowski, O. Shehory (Eds.), Cooperative Information Agents VI. Proceedings, 2002. XI, 321 pages. 2002. (Subseries LNAI).

Vol. 2447: D.J. Hand, N.M. Adams, R.J. Bolton (Eds.), Pattern Detection and Discovery. Proceedings, 2002. XII, 227 pages. 2002. (Subseries LNAI).

Vol. 2448: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. Proceedings, 2002. XII, 481 pages. 2002. (Subseries LNAI).

Vol. 2449: L. Van Gool (Ed.), Pattern Recognotion. Proceedings, 2002. XVI, 628 pages. 2002.

Vol. 2451: B. Hochet, A.J. Acosta, M.J. Bellido (Eds.), Integrated Circuit Design. Proceedings, 2002. XVI, 496 pages. 2002.

Vol. 2452: R. Guigó, D. Gusfield (Eds.), Algorithms in Bioinformatics. Proceedings, 2002. X, 554 pages. 2002.

Vol. 2453: A. Hameurlain, R. Cicchetti, R. Traunmüller (Eds.), Database and Expert Systems Applications. Proceedings, 2002. XVIII, 951 pages. 2002.

Vol. 2454: Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.), Data Warehousing and Knowledge Discovery. Proceedings, 2002. XIII, 339 pages. 2002.

Vol. 2455: K. Bauknecht, A M. Tjoa, G. Quirchmayr (Eds.), E-Commerce and Web Technologies. Proceedings, 2002. XIV, 414 pages. 2002.

Vol. 2456: R. Traunmüller, K. Lenk (Eds.), Electronic Government. Proceedings, 2002. XIII, 486 pages. 2002.

Vol. 2458: M. Agosti, C. Thanos (Eds.), Research and Advanced Technology for Digital Libraries. Proceedings, 2002. XVI, 664 pages. 2002.

Vol. 2462: K. Jansen, S. Leonardi, V. Vazirani (Eds.), Approximation Algorithms for Combinatorial Optimization. Proceedings, 2002. VIII, 271 pages. 2002.

Vol. 2463: M. Dorigo, G. Di Caro, M. Sampels (Eds.), Ant Algorithms. Proceedings, 2002. XIII, 305 pages. 2002.

Vol. 2464: M. O'Neill, R.F.E. Sutcliffe, C. Ryan, M. Eaton, N. Griffith (Eds.), Artificial Intelligence and Cognitive Science. Proceedings, 2002. XI, 247 pages. 2002. (Subseries LNAI).

Vol. 2469: W. Damm, E.-R. Olderog (Eds.), Formal Techniques in Real-Time and Fault-Tolerant Systems. Proceedings, 2002. X, 455 pages. 2002.

Vol. 2470: P. Van Hentenryck (Ed.), Principles and Practice of Constraint Programming – CP 2002. Proceedings, 2002. XVI, 794 pages. 2002.

Vol. 2479: M. Jarke, J. Koehler, G. Lakemeyer (Eds.), KI 2002: Advances in Artificial Intelligence. Proceedings, 2002. XIII, 327 pages. (Subseries LNAI).

Vol. 2483: J.D.P. Rolim, S. Vadhan (Eds.), Randomization and Approximation Techniques in Computer Science. Proceedings, 2002. VIII, 275 pages. 2002.

Table of Contents

Regular Papers

On the Usefulness of Extracting Syntactic Dependencies for Text Indexing 3
Miguel A. Alonso, Jesús Vilares, Víctor M. Darriba

Using Latent Semantic Indexing as a Measure of Conceptual Association for Noun Compound Disambiguation 12
Alan M. Buckeridge, Richard F.E. Sutcliffe

RADAR: Finding Analogies Using Attributes of Structure 20
Brian P. Crean, Diarmuid O'Donoghue

Classifying Languages Based on Speech Rhythm 28
Fred Cummins

Finding Agents in a Two-Dimensional Boolean STaM 36
Jim Doran

Neuro-symbolic System for Forecasting Red Tides 45
Florentino Fdez-Riverola, Juan M. Corchado, Jesús M. Torres

Improved Learning for Hidden Markov Models Using Penalized Training 53
Bill Keller, Rudi Lutz

Recovering High-Level Structure of Software Systems Using a Minimum Description Length Principle 61
Rudi Lutz

A System for Multi-agent Information Retrieval 70
Paula Mc Dermott, Colm O'Riordan

All There Is to the Mind Is to Have the Right Genes, or, Consciousness as a Form of Genetic Engineering 78
Ajit Narayanan

Towards Robust Collaborative Filtering 87
Michael P. O'Mahony, Neil J. Hurley, Guenole C.M. Silvestre

GVR: A New Genetic Representation for the Vehicle Routing Problem ... 95
Francisco B. Pereira, Jorge Tavares, Penousal Machado, Ernesto Costa

An Empirical Comparison of Particle Swarm and Predator Prey Optimisation 103
Arlindo Silva, Ana Neves, Ernesto Costa

Data Mining Support for Case-Based Collaborative Recommendation 111
Barry Smyth, David Wilson, Derry O'Sullivan

The Feasibility of Machine Learning for Query Answering – An Experiment in Two Domains 119
Richard F.E. Sutcliffe, Kieran White

Meta-knowledge Annotation for Efficient Natural-Language Question-Answering 127
Tony Veale

Concise Papers

Financial Time Series Modelling Using Neural Networks: An Assessment of the Utility of a Stacking Methodology 137
Anthony Brabazon

Experiments in Sparsity Reduction: Using Clustering in Collaborative Recommenders 144
Derek Bridge, Jerome Kelleher

Identification of Visual Features Using a Neural Version of Exploratory Projection Pursuit 150
Emilio Corchado, Colin Fyfe

How People Compare an Item's Placement in Two Alternative Categories 158
Fintan Costello

Investigations into Market Index Trading Models Using Evolutionary Automatic Programming 165
Ian Dempsey, Michael O'Neill, Anthony Brabazon

An Interactive Story Engine 171
Chris Fairclough, Pádraig Cunningham

Coherence, Explanation, and Bayesian Networks 177
David H. Glass

Combining Case-Based Reasoning and Analogical Reasoning in Software Design 183
Paulo Gomes, Francisco C. Pereira, Nuno Seco, Paulo Paiva, Paulo Carreiro, José L. Ferreira, Carlos Bento

Combination Methods for Improving the Reliability of Machine Translation Based Cross-Language Information Retrieval 190
Gareth J.F. Jones, Adenike M. Lam-Adesina

A System for Music Information Retrieval 197
Orla Lahart, Colm O'Riordan

A New Bayesian Network Structure for Classification Tasks 203
Michael G. Madden

Evaluating Preference-Based Feedback in Recommender Systems 209
Lorraine Mc Ginty, Barry Smyth

Design of a Musical Instrument Classifier System Based on Mel Scaled Cepstral Coefficient Supervectors and a Supervised Two-Layer Feedforward Neural Network 215
Declan McGrath

An Interactive Learning Environment for Knowledge Engineering 221
David McSherry

Customising a Copying-Identifier for Biomedical Science Student Reports: Comparing Simple and Smart Analyses 228
Julia Medori, Eric Atwell, Paul Gent, Clive Souter

A Hybridised GA for the Steiner Minimal Tree Problem 234
R. Panadero, J.L. Fernández-Villacañas

Speaking Autonomous Intelligent Devices 240
Robert J. Ross, Bryan McEleney, Robert Kelly, Tarek Abu-Amer, Michael Walsh, Julie Carson-Berndsen, Gregory M.P. O'Hare

Author Index 247

Regular Papers

On the Usefulness of Extracting Syntactic Dependencies for Text Indexing*

Miguel A. Alonso[1], Jesús Vilares[1,2], and Víctor M. Darriba[2]

[1] Departamento de Computación, Universidade da Coruña
Campus de Elviña s/n, 15071 La Coruña, Spain
alonso@udc.es, jvilares@mail2.udc.es
http://coleweb.dc.fi.udc.es/

[2] Escuela Superior de Ingeniería Informática, Universidade de Vigo
Campus de As Lagoas, 32004 Orense, Spain
jvilares@uvigo.es, darriba@ei.uvigo.es

Abstract. In recent years, there has been a considerable amount of interest in using Natural Language Processing in Information Retrieval research, with specific implementations varying from the word-level morphological analysis to syntactic parsing to conceptual-level semantic analysis. In particular, different degrees of phrase-level syntactic information have been incorporated in information retrieval systems working on English or Germanic languages such as Dutch. In this paper we study the impact of using such information, in the form of syntactic dependency pairs, in the performance of a text retrieval system for a Romance language, Spanish.

1 Introduction

For Information Retrieval (IR) tasks, documents are frequently represented through a set of index terms or representative keywords. This can be accomplished through operations such as the elimination of *stopwords* (too frequent words or words with no apparent significance) or the use of *stemming* (which reduces distinct words to their supposed grammatical root). These operations are called *text operations*, providing a *logical view* of the processed document. More elaborated index terms can be created by combining two or more content words (nouns, verbs and adjectives) in a *multi-word term* [11,6,12]. Most techniques for extracting multi-word terms rely on statistics [7] or simple pattern matching [12], instead of considering the structural relations among the words that form a sentence. In this paper, we propose to use practical, finite-state, Natural Language Processing (NLP) techniques to extract such multi-word terms in the form of pairs of words related by some kind of syntactic dependency.

* The research reported in this article has been supported in part by Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (Grant TIC2000-0370-C02-01), Ministerio de Ciencia y Tecnología (Grant HP2001-0044) and Xunta de Galicia (Grant PGIDT01PXI10506PN).

M. O'Neill et al. (Eds.): AICS 2002, LNAI 2464, pp. 3–11, 2002.

2 Extraction of Syntactic Dependencies

Given a stream of tagged words, we want to obtain the *head-modifier* pairs corresponding to the most relevant syntactic dependencies [5]: *Noun-Modifier*, relating the head of a noun phrase with the head of a modifier; *Subject-Verb*, relating the head of the subject with the main verb of the clause; and *Verb-Complement*, relating the main verb of the clause with the head of a complement. It has to be noted that while the head-modifier relation may suggest semantic dependence, what we obtain here is strictly syntactic, even though the semantic relation is what we are really after [16].

The kernel of the grammar used by our shallow parser has been inferred from the basic trees corresponding to noun phrases and their syntactic and morpho-syntactic variants [11]. *Syntactic variants* result from the inflection of individual words and from modifying the syntactic structure of the original noun phrase. *Morpho-syntactic variants* differ from syntactic variants in that at least one of the content words of the original noun phrase is transformed into another word derived from the same morphological stem. At this point we must recall that inflectional morphemes represent grammatical concepts such as gender, person, mood, or time and tense. On the other hand, derivational morphemes effect a semantic change on the base, often also effecting a change of syntactic class. We define a *morphological family* as the set of words obtained from the same morphological root through derivation mechanisms, such as prefixation, emotive suffixation, non-emotive suffixation, back formation and parasynthesis. A system for the automatic generation of morphological families has been described in [18].

2.1 Syntactic Variants

The example of Fig. 1 shows the basic structure of a noun phrase and some of its possible syntactic variants, together with the syntactic dependencies they contain. Such variants were obtained by applying to the source phrase *una caída de las ventas* (a drop in the sales) the following mechanisms [11]:

- *Synapsy:* a unary construction which corresponds to a change of preposition or the addition or removal of a determiner.
- *Substitution:* it consists of employing modifiers to make a term more specific.
- *Permutation:* this refers to the permutation of words around a pivot element.
- *Coordination:* this consists of employing coordinating constructions (copulative or disjunctive) with the modifier or with the modified term.

Symbols A, C, D, N, P, V and W are the part-of-speech labels that denote adjectives, coordinating conjunctions, determiners, nouns, prepositions, verbs and adverbs, respectively. In addition, we have conflated each word in a dependency pair by replacing it with an identifier of its morphological family (actually, one of the words in such family, its *representative*).

The structures and syntactic dependencies corresponding to all the syntactic variants shown in Fig. 1 are embedded in the syntactic pattern shown in Fig. 2