Marcin Detyniecki Joemon M. Jose
Andreas Nürnberger C.J. van Rijsbergen (Eds.)

LNCS 3877

Adaptive Multimedia Retrieval: User, Context, and Feedback

Third International Workshop, AMR 2005
Glasgow, UK, July 2005
Revised Selected Papers

Springer

Marcin Detyniecki
Joemon M. Jose
Andreas Nürnberger
C.J. van Rijsbergen (Eds.)

Adaptive Multimedia Retrieval: User, Context, and Feedback

Third International Workshop, AMR 2005
Glasgow, UK, July 28-29, 2005
Revised Selected Papers

Springer

Volume Editors

Marcin Detyniecki
Laboratoire d'Informatique de Paris 6, LIP6
8 rue du Capitaine Scott, 75015 Paris, France
E-mail: Marcin.Detyniecki@lip6.fr

Joemon M. Jose
C.J. van Rijsbergen
University of Glasgow
Department of Computing Science
University Avenue, Glasgow G12 8QQ, UK
E-mail: {jj,keith}@dcs.gla.ac.uk

Andreas Nürnberger
Otto-von-Guericke Universität Madgeburg
Fakultät für Informatik
Universitätsplatz 2, 39106 Magdeburg, Germany
E-mail: nuernb@iws.cs.uni-magdeburg.de

Library of Congress Control Number: 2006920781

CR Subject Classification (1998): H.3, H.5.1, H.5.5, I.4, I.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-32174-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32174-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11670834 06/3142 5 4 3 2 1 0

Preface

This book is an extended collection of revised contributions that were initially submitted to the International Workshop on Adaptive Multimedia Retrieval (AMR 2005). This workshop was organized during July 28-29, 2005, at the University of Glasgow, UK, as part of an information retrieval research festival and in co-location with the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005). AMR 2005 was the third and so far the biggest event of the series of workshops that started in 2003 with a workshop during the 26th German Conference on Artificial Intelligence (KI 2003) and continued in 2004 as part of the 16th European Conference on Artificial Intelligence (ECAI 2004).

The workshop focussed especially on intelligent methods to analyze and structure multimedia collections, with particular attention on methods that are able to support the user in the search process, e.g., by providing additional user-and context-adapted information about the search results as well as the data collection itself and especially by adapting the retrieval tool to the user's needs and interests. The invited contributions presented in the first section of this book—"Putting the User in the Loop: Visual Resource Discovery" from Stefan Rüger, "Using Relevance Feedback to Bridge the Semantic Gap" from Ebroul Izquierdo and Divna Djordjevic, and "Leveraging Context for Adaptive Multimedia Retrieval: A Matter of Control" from Gary Marchionini—illustrate these core topics: user, context and feedback. These aspects are discussed from different points of view in the 18 contributions that are classified into six main chapters, following rather closely the workshop's sessions: ranking, systems, spatio-temporal relations, using feedback, using context and meta-data. We think that this book provides a good and conclusive overview of the current research in this area.

We would like to thank all members of the Program Committee for supporting us in the reviewing process, the workshop participants for their willingness to revise and extend their papers for this book and Alfred Hofmann from Springer for his support in publishing this book.

October 2005

Marcin Detyniecki, Paris, France
Joemon M. Jose, Glasgow, UK
Andreas Nürnberger, Magdeburg, Germany
Keith van Rijsbergen, Glasgow, UK

Sponsoring Institutions

MMKM – Multimedia Knowledge Management Network, UK
University of Glasgow, UK
BCS – The British Computer Society, UK
British HCI Group, UK
UK Sharp Laboratories of Europe Ltd.

Organization

General Chair

Keith van Rijsbergen	University of Glasgow, UK

Program Chairs

Andreas Nürnberger	University of Magdeburg, Germany
Marcin Detyniecki	CNRS, Lab. d'Informatique de Paris 6, France
Joemon M. Jose	University of Glasgow, UK

Local Chair

Iadh Ounis	University of Glasgow, UK

Publicity Chair

Jana Urban	University of Glasgow, UK

Program Committee

Jenny Benois-Pineau	University of Bordeaux, LABRI, France
Pia Borlund	Royal School of Libr. and Inform. Science, Denmark
Arjen De Vries	CWI, Amsterdam, The Netherlands
Norbert Fuhr	University of Duisburg-Essen, Germany
Bogdan Gabrys	Bournemouth University, UK
Ana M. García Serrano	Universidad Politécnica de Madrid, Spain
Sebastian Goeser	IBM Germany Development Ltd., Germany
Philippe Joly	Université Paul Sabatier, Toulouse, France
Gareth Jones	Dublin City University, Ireland
Pietro Pala	Università di Firenze, Italy
Stefanos Kollias	National Technical University of Athens, Greece
Stéphane Marchand-Maillet	University of Geneva, Switzerland
Trevor Martin	University of Bristol, UK
José María Martínez Sánchez	Universidad Autónoma de Madrid, Spain
Bernard Merialdo	Sophia Antipolis Cédex, France
Gheorghe Muresan	Rutgers University, USA
Stefan Rüger	Imperial College London, UK

Nicu Sebe	Leiden University, The Netherlands
Ingo Schmitt	University of Magdeburg, Germany
Alan F. Smeaton	Dublin City University, Ireland

Lecture Notes in Computer Science

For information about Vols. 1–3781

please contact your bookseller or Springer

Vol. 3884: B. Durand, W. Thomas (Eds.), STACS 2006. XIV, 714 pages. 2006.

Vol. 3879: T. Erlebach, G. Persinao (Eds.), Approximation and Online Algorithms. X, 349 pages. 2006.

Vol. 3878: A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing. XVII, 589 pages. 2006.

Vol. 3877: M. Detyniecki, J.M. Jose, A. Nürnberger, C. J. van Rijsbergen (Eds.), Adaptive Multimedia Retrieval: User, Context, and Feedback. XI, 279 pages. 2006.

Vol. 3874: R. Missaoui, J. Schmidt (Eds.), Formal Concept Analysis. X, 309 pages. 2006. (Sublibrary LNAI).

Vol. 3872: H. Bunke, A. L. Spitz (Eds.), Document Analysis Systems VII. XIII, 630 pages. 2006.

Vol. 3870: S. Spaccapietra, P. Atzeni, W.W. Chu, T. Catarci, K.P. Sycara (Eds.), Journal on Data Semantics V. XIII, 237 pages. 2006.

Vol. 3869: S. Renals, S. Bengio (Eds.), Machine Learning for Multimodal Interaction. XIII, 490 pages. 2006.

Vol. 3868: K. Römer, H. Karl, F. Mattern (Eds.), Wireless Sensor Networks. XI, 342 pages. 2006.

Vol. 3863: M. Kohlhase (Ed.), Mathematical Knowledge Management. XI, 405 pages. 2006. (Sublibrary LNAI).

Vol. 3861: J. Dix, S.J. Hegner (Eds.), Foundations of Information and Knowledge Systems. X, 331 pages. 2006.

Vol. 3860: D. Pointcheval (Ed.), Topics in Cryptology – CT-RSA 2006. XI, 365 pages. 2006.

Vol. 3858: A. Valdes, D. Zamboni (Eds.), Recent Advances in Intrusion Detection. X, 351 pages. 2006.

Vol. 3857: M.P.C. Fossorier, H. Imai, S. Lin, A. Poli (Eds.), Applied Algebra, Algebraic Algorithms and Error-Correcting Codes. XI, 350 pages. 2006.

Vol. 3855: E. A. Emerson, K.S. Namjoshi (Eds.), Verification, Model Checking, and Abstract Interpretation. XI, 443 pages. 2005.

Vol. 3853: A.J. Ijspeert, T. Masuzawa, S. Kusumoto (Eds.), Biologically Inspired Approaches to Advanced Information Technology. XIV, 388 pages. 2006.

Vol. 3852: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision - ACCV 2006, Part II. XXXI, 977 pages. 2005.

Vol. 3851: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision - ACCV 2006, Part I. XXXI, 973 pages. 2006.

Vol. 3850: R. Freund, G. Păun, G. Rozenberg, A. Salomaa (Eds.), Membrane Computing. IX, 371 pages. 2006.

Vol. 3848: J.-F. Boulicaut, L. De Raedt, H. Mannila (Eds.), Constraint-Based Mining and Inductive Databases. X, 401 pages. 2006. (Sublibrary LNAI).

Vol. 3847: K.P. Jantke, A. Lunzer, N. Spyratos, Y. Tanaka (Eds.), Federation over the Web. X, 215 pages. 2006. (Sublibrary LNAI).

Vol. 3846: H. J. van den Herik, Y. Björnsson, N.S. Netanyahu (Eds.), Computers and Games. XIV, 333 pages. 2006.

Vol. 3845: J. Farré, I. Litovsky, S. Schmitz (Eds.), Implementation and Application of Automata. XIII, 360 pages. 2006.

Vol. 3844: J.-M. Bruel (Ed.), Satellite Events at the MoDELS 2005 Conference. XIII, 360 pages. 2006.

Vol. 3843: P. Healy, N.S. Nikolov (Eds.), Graph Drawing. XVII, 536 pages. 2006.

Vol. 3842: H.T. Shen, J. Li, M. Li, J. Ni, W. Wang (Eds.), Advanced Web and Network Technologies, and Applications. XXVII, 1057 pages. 2006.

Vol. 3841: X. Zhou, J. Li, H.T. Shen, M. Kitsuregawa, Y. Zhang (Eds.), Frontiers of WWW Research and Development - APWeb 2006. XXIV, 1223 pages. 2006.

Vol. 3840: M. Li, B. Boehm, L.J. Osterweil (Eds.), Unifying the Software Process Spectrum. XVI, 522 pages. 2006.

Vol. 3839: J.-C. Filliâtre, C. Paulin-Mohring, B. Werner (Eds.), Types for Proofs and Programs. VIII, 275 pages. 2006.

Vol. 3838: A. Middeldorp, V. van Oostrom, F. van Raamsdonk, R. de Vrijer (Eds.), Processes, Terms and Cycles: Steps on the Road to Infinity. XVIII, 639 pages. 2005.

Vol. 3837: K. Cho, P. Jacquet (Eds.), Technologies for Advanced Heterogeneous Networks. IX, 307 pages. 2005.

Vol. 3836: J.-M. Pierson (Ed.), Data Management in Grids. X, 143 pages. 2006.

Vol. 3835: G. Sutcliffe, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XIV, 744 pages. 2005. (Sublibrary LNAI).

Vol. 3834: D.G. Feitelson, E. Frachtenberg, L. Rudolph, U. Schwiegelshohn (Eds.), Job Scheduling Strategies for Parallel Processing. VIII, 283 pages. 2005.

Vol. 3833: K.-J. Li, C. Vangenot (Eds.), Web and Wireless Geographical Information Systems. XI, 309 pages. 2005.

Vol. 3832: D. Zhang, A.K. Jain (Eds.), Advances in Biometrics. XX, 796 pages. 2005.

Vol. 3831: J. Wiedermann, G. Tel, J. Pokorný, M. Bieliková, J. Štuller (Eds.), SOFSEM 2006: Theory and Practice of Computer Science. XV, 576 pages. 2006.

Vol. 3829: P. Pettersson, W. Yi (Eds.), Formal Modeling and Analysis of Timed Systems. IX, 305 pages. 2005.

Vol. 3828: X. Deng, Y. Ye (Eds.), Internet and Network Economics. XVII, 1106 pages. 2005.

Vol. 3827: X. Deng, D.-Z. Du (Eds.), Algorithms and Computation. XX, 1190 pages. 2005.

Vol. 3826: B. Benatallah, F. Casati, P. Traverso (Eds.), Service-Oriented Computing - ICSOC 2005. XVIII, 597 pages. 2005.

Vol. 3824: L.T. Yang, M. Amamiya, Z. Liu, M. Guo, F.J. Rammig (Eds.), Embedded and Ubiquitous Computing – EUC 2005. XXIII, 1204 pages. 2005.

Vol. 3823: T. Enokido, L. Yan, B. Xiao, D. Kim, Y. Dai, L.T. Yang (Eds.), Embedded and Ubiquitous Computing – EUC 2005 Workshops. XXXII, 1317 pages. 2005.

Vol. 3822: D. Feng, D. Lin, M. Yung (Eds.), Information Security and Cryptology. XII, 420 pages. 2005.

Vol. 3821: R. Ramanujam, S. Sen (Eds.), FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science. XIV, 566 pages. 2005.

Vol. 3820: L.T. Yang, X.-s. Zhou, W. Zhao, Z. Wu, Y. Zhu, M. Lin (Eds.), Embedded Software and Systems. XXVIII, 779 pages. 2005.

Vol. 3819: P. Van Hentenryck (Ed.), Practical Aspects of Declarative Languages. X, 231 pages. 2005.

Vol. 3818: S. Grumbach, L. Sui, V. Vianu (Eds.), Advances in Computer Science – ASIAN 2005. XIII, 294 pages. 2005.

Vol. 3817: M. Faundez-Zanuy, L. Janer, A. Esposito, A. Satue-Villar, J. Roure, V. Espinosa-Duro (Eds.), Nonlinear Analyses and Algorithms for Speech Processing. XII, 380 pages. 2006. (Sublibrary LNAI).

Vol. 3816: G. Chakraborty (Ed.), Distributed Computing and Internet Technology. XXI, 606 pages. 2005.

Vol. 3815: E.A. Fox, E.J. Neuhold, P. Premsmit, V. Wuwongse (Eds.), Digital Libraries: Implementing Strategies and Sharing Experiences. XVII, 529 pages. 2005.

Vol. 3814: M. Maybury, O. Stock, W. Wahlster (Eds.), Intelligent Technologies for Interactive Entertainment. XV, 342 pages. 2005. (Sublibrary LNAI).

Vol. 3813: R. Molva, G. Tsudik, D. Westhoff (Eds.), Security and Privacy in Ad-hoc and Sensor Networks. VIII, 219 pages. 2005.

Vol. 3811: C. Bussler, M.-C. Shan (Eds.), Technologies for E-Services. VIII, 127 pages. 2006.

Vol. 3810: Y.G. Desmedt, H. Wang, Y. Mu, Y. Li (Eds.), Cryptology and Network Security. XI, 349 pages. 2005.

Vol. 3809: S. Zhang, R. Jarvis (Eds.), AI 2005: Advances in Artificial Intelligence. XXVII, 1344 pages. 2005. (Sublibrary LNAI).

Vol. 3808: C. Bento, A. Cardoso, G. Dias (Eds.), Progress in Artificial Intelligence. XVIII, 704 pages. 2005. (Sublibrary LNAI).

Vol. 3807: M. Dean, Y. Guo, W. Jun, R. Kaschek, S. Krishnaswamy, Z. Pan, Q.Z. Sheng (Eds.), Web Information Systems Engineering – WISE 2005 Workshops. XV, 275 pages. 2005.

Vol. 3806: A.H. H. Ngu, M. Kitsuregawa, E.J. Neuhold, J.-Y. Chung, Q.Z. Sheng (Eds.), Web Information Systems Engineering – WISE 2005. XXI, 771 pages. 2005.

Vol. 3805: G. Subsol (Ed.), Virtual Storytelling. XII, 289 pages. 2005.

Vol. 3804: G. Bebis, R. Boyle, D. Koracin, B. Parvin (Eds.), Advances in Visual Computing. XX, 755 pages. 2005.

Vol. 3803: S. Jajodia, C. Mazumdar (Eds.), Information Systems Security. XI, 342 pages. 2005.

Vol. 3802: Y. Hao, J. Liu, Y.-P. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma, Y.-C. Jiao (Eds.), Computational Intelligence and Security, Part II. XLII, 1166 pages. 2005. (Sublibrary LNAI).

Vol. 3801: Y. Hao, J. Liu, Y.-P. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma, Y.-C. Jiao (Eds.), Computational Intelligence and Security, Part I. XLI, 1122 pages. 2005. (Sublibrary LNAI).

Vol. 3799: M. A. Rodríguez, I.F. Cruz, S. Levashkin, M.J. Egenhofer (Eds.), GeoSpatial Semantics. X, 259 pages. 2005.

Vol. 3798: A. Dearle, S. Eisenbach (Eds.), Component Deployment. X, 197 pages. 2005.

Vol. 3797: S. Maitra, C. E. V. Madhavan, R. Venkatesan (Eds.), Progress in Cryptology - INDOCRYPT 2005. XIV, 417 pages. 2005.

Vol. 3796: N.P. Smart (Ed.), Cryptography and Coding. XI, 461 pages. 2005.

Vol. 3795: H. Zhuge, G.C. Fox (Eds.), Grid and Cooperative Computing - GCC 2005. XXI, 1203 pages. 2005.

Vol. 3794: X. Jia, J. Wu, Y. He (Eds.), Mobile Ad-hoc and Sensor Networks. XX, 1136 pages. 2005.

Vol. 3793: T. Conte, N. Navarro, W.-m.W. Hwu, M. Valero, T. Ungerer (Eds.), High Performance Embedded Architectures and Compilers. XIII, 317 pages. 2005.

Vol. 3792: I. Richardson, P. Abrahamsson, R. Messnarz (Eds.), Software Process Improvement. VIII, 215 pages. 2005.

Vol. 3791: A. Adi, S. Stoutenburg, S. Tabet (Eds.), Rules and Rule Markup Languages for the Semantic Web. X, 225 pages. 2005.

Vol. 3790: G. Alonso (Ed.), Middleware 2005. XIII, 443 pages. 2005.

Vol. 3789: A. Gelbukh, Á. de Albornoz, H. Terashima-Marín (Eds.), MICAI 2005: Advances in Artificial Intelligence. XXVI, 1198 pages. 2005. (Sublibrary LNAI).

Vol. 3788: B. Roy (Ed.), Advances in Cryptology - ASIACRYPT 2005. XIV, 703 pages. 2005.

Vol. 3787: D. Kratsch (Ed.), Graph-Theoretic Concepts in Computer Science. XIV, 470 pages. 2005.

Vol. 3786: J. Song, T. Kwon, M. Yung (Eds.), Information Security Applications. XI, 378 pages. 2006.

Vol. 3785: K.-K. Lau, R. Banach (Eds.), Formal Methods and Software Engineering. XIV, 496 pages. 2005.

Vol. 3784: J. Tao, T. Tan, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction. XIX, 1008 pages. 2005.

Vol. 3783: S. Qing, W. Mao, J. Lopez, G. Wang (Eds.), Information and Communications Security. XIV, 492 pages. 2005.

Vol. 3782: K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, T.R. Roth-Berghofer (Eds.), Professional Knowledge Management. XXIII, 739 pages. 2005. (Sublibrary LNAI).

Table of Contents

Invited Contributions

Ranking

Systems

Spatio-temporal Relations

Using Feedback

Using Context

Meta Data

Putting the User in the Loop: Visual Resource Discovery

Stefan Rüger

Department of Computing, South Kensington Campus,
Imperial College London, London SW7 2AZ, UK
s.rueger@imperial.ac.uk

Abstract. Visual resource discovery modes are discussed with a view to apply them in a wide variety of digital multimedia collections. The paradigms include summarising complex multimedia objects such as TV news, information visualisation techniques for document clusters, visual search by example, relevance feedback and methods to create browsable structures within the collection. These exploration modes share three common features: they are automatically generated, depend on visual senses and interact with the user of the multimedia collections.

1 Introduction

Giving users access to collections is one of the defining tasks of a library. For thousands of years the traditional methods of resource discovery have been searching, browsing and asking: Librarians create reference cards with meta-data that are put into catalogues (nowadays databases); they also place the objects in physical locations that follow certain classification schemes and they answer questions at the reference desk.

The advent of digital documents has radically changed the organisation principles: Now it is possible to *automatically* index and search document collections as big as the world-wide web *à la* Google and browse collections utilising author-inserted links. It is almost as if automated processing has turned the traditional library access upside down: instead of searching meta-data catalogues in order to retrieve the document, web search engines search the full content of documents and retrieve their meta-data, ie, the location where documents can be found. Undoubtedly, this automated approach has made all the difference to the way the vast world-wide web can be utilised.

However, indexing sheer mass is no guarantee of success either: While most of today's inter-library loan systems allow access to virtually any publication in the world (at least to around 40m entries in OCLC's Worldcat database and a further 3m from Bowker's Books In Print), students and researchers alike seem to be reluctant to actually make use of this facility. On the other hand, the much smaller catalogue offered by Amazon appears to be very popular — presumably owing to added services such as subject categories; fault tolerant search tools; personalised services telling the customer what's new in a subject area or what

M. Detyniecki et al. (Eds.): AMR 2005, LNCS 3877, pp. 1–18, 2006.

other people with a similar profile bought; pictures of book covers; media and customer reviews; access to the table of contents, to selections of the text and to the full-text index of popular books; and the perception of fast delivery.

This paper argues that automated added services such as visual queries, browsing and summaries can prove useful for resource discovery in multimedia digital libraries. Multimedia collections pose their very own challenges in this context; images and videos don't usually come with dedicated reference cards or meta-data, and when they do, as in museum collections, their creation will have been expensive and time-consuming. The next section explores methods of automatically indexing, labelling and annotating image and video content. It briefly discusses the challenges of the semantic gap, polysemy, fusion and responsiveness inherent with these. Sections 3 and 4 are about summarising techniques for videos and about visualisation of search results, while Section 5 discusses content-based visual search modes such as query by example and relevance feedback. Section 6 promotes browsing as resource discovery mode and looks at underlying techniques to automatically structure the document collection.

2 Challenges of Automated Visual Indexing

Videos can be annotated, ie, get indexable text strings assigned, using a variety of sources: closed-captions, teletext, subtitles, automated speech recognition on the audio and optical character recognition for text embedded in the frames of a video. The resulting text strings are then used as the basis for full-text indexing, which is the way most video retrieval systems operate, including Google's latest TV search engine http://video.google.com.

Automatically annotating images with text strings is less straightforward. Methods attempting this task include dedicated machine vision models for particular words (such as 'people' or 'aeroplane'); machine translation methods that link image regions (blobs) and words in the same way as corresponding words in two text documents written in different languages but otherwise of same contents [30]; co-occurrence models of low-level image features of tiled image regions and words [52]; cross-lingual information retrieval models [46, 48]; inference networks that connect image segments with words [51]; probabilistic modelling with Latent Dirlichet Allocation [9], Bernoulli distributions [31] or non-parametric density estimation [81]; Support Vector Machine classification and relevance feedback [45]; and simple scene-level statistics [72]. All these methods have in common that a controlled vocabulary set of limited size (in the order of 500 more or less general terms) is used to annotate images based on a large training set.

The commonest way of indexing the visual content of images is by extracting low-level features, which represent colour usage, texture composition, shape and structure, localisation or motion. These representations are often real-valued vectors containing summary statistics, eg, in the form of histograms; their respective distances act as indicators whether or not two images are similar with respect to this particular feature. Design and usage of these features can be critical, and there is a wide variety of them, eg, as published by participants in the TRECVID conference [74, 34]. Once created, those features will allow the comparison and ranking of images in the database with respect to images submitted as a query (*query-by-example*).

There are a number of open issues with this approach: On a perceptual level, those low-level features do not necessarily correlate with any high-level meaning the images might have, such as victory or triumph. Even if they did, images usually convey a multitude of meanings so that the query-by-example approach is bound to under-specify the real information need. The former problem is known as *semantic gap* and the latter as *polysemy*. Designing a human-computer interaction that utilises the user's feedback has been one of the main approaches to tackle these perceptual issues. Amongst other methods there are those that seek to reformulate the query [44,49,58] or those that weight the various features differently depending on the user's feedback. Weight adaptation methods include non-parametric density estimation [50]; cluster analysis of the images [79]; transposed files for feature selection [71]; Bayesian network learning [21,50]; statistical analysis of the feature distributions of relevant images [60]; variance analysis [60]; and analytic global optimisation [36,40]. Some approaches give the presentation and placement of images on screen much consideration to indicate similarity of images amongst themselves [64, 59] or with respect to a visual query [36, 40].

On a practical level, the multitude of features assigned to images poses a *fusion problem:* how to combine possibly conflicting evidence of two images' similarity? There are many approaches to carry out fusion, some based on labelled training data and some based on user feedback for the current query [3, 7, 67, 80].

There is a *responsiveness problem*, too, in that the naïve comparison of query feature vectors to the database feature vectors requires a linear scan through the database. Although the scan is eminently scalable, the practicalities of doing this operation can mean an undesirable response time in the order of seconds rather than the 100 milli-seconds that can be achieved by text search engines. The problem is that high-dimensional tree structures tend to collapse to linear scans above a certain dimensionality [77]. As a consequence, some approaches for fast nearest-neighbour search use compression techniques to speed up the disk access of linear scan as in [77] using VA-files; or they approximate the search [54, 8]; decompose the features componentwise [28, 1, 14] saving access to unnecessary components; or deploy a combination of these [53, 43].

3 Video Summaries

Even if the automated methods of the preceding section enabled a retrieval process with high precision (proportion of the retrieved items that are relevant) and high recall (proportion of the relevant items that are retrieved) it would still be vital to present the retrieval results in a way so that the users can quickly decide whether or not those items are relevant to them.

Images are most naturally displayed as thumbnails, and their relevance can quickly be judged by users. Presenting and summarising videos is a bit more involved. The main metaphor used for this is that of a *storyboard* that would contain *keyframes* with some text about the video. Several systems exist that summarise news stories in this way, most notably Informedia [18] and Físchlár [69]. The Informedia system devotes much effort to added services such as face recognition and speaker voice identification allowing retrieval of the appearance of known people. Informedia also provides alternative modes of presentation, eg, through film skims or by assembling 'collages' of images, text and other information (eg, maps) sourced via references from the text [17]. Físchlár's added value lies in the ability to personalise the content (with the user expressing like or dislike of stories) and in assembling lists of related stories and recommendations.

Our very own TV news search engine ANSES [57, 56] records the main BBC evening news along with the sub-titles, indexes them, breaks the video stream

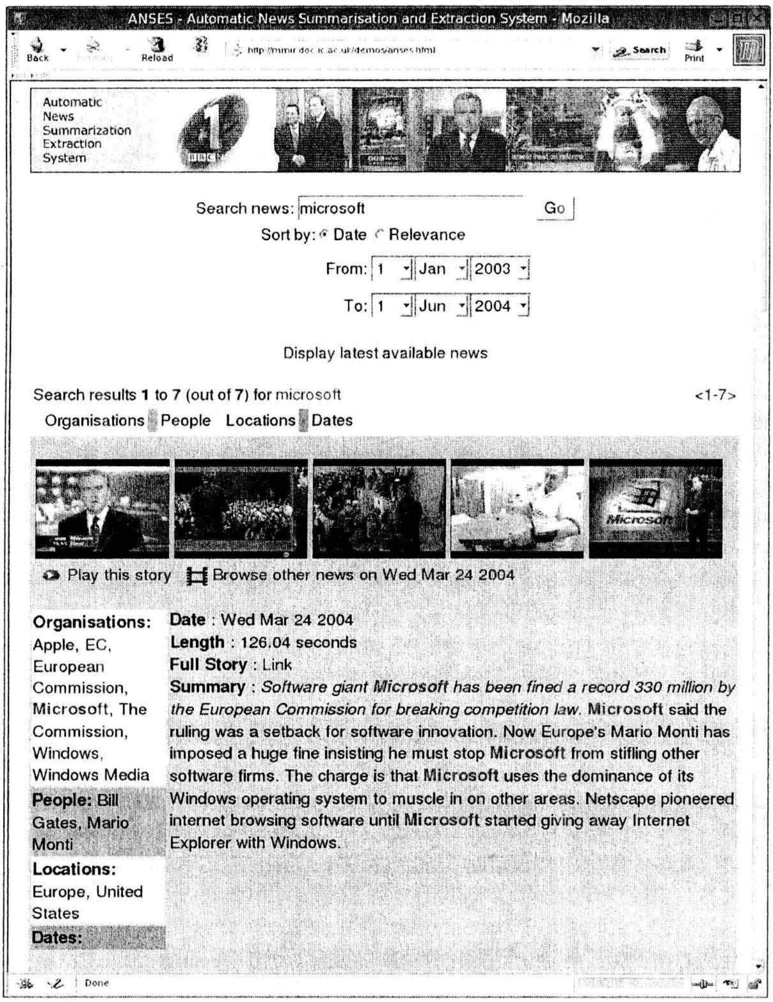

Fig. 1. News search engine interface

into shots (defined as those video sequences that are generated during a continuous operation of the camera), extracts one key-frame per shot, automatically glues shots together to form news stories based on an overlap in vocabulary in the sub-titles of adjacent shots (using lexical chains), and assembles a story-board for each story that can be retrieved using text searches or browsed. Fig 1 shows the interface of ANSES. The natural language toolset GATE [26, 27] is used for automated discovery of organisations, people, places and dates to provide an instant indication of what the news story is about. ANSES also displays a short automated textual extraction summary, again using lexical chains to identify the most salient sentences. These summaries are never as informative as hand-made ones, but users of the system have found them crucial for judging whether or not they are interested in a particular returned search result.

Dissecting the video stream into shots and associating one keyframe along with text from subtitles to each shot has another advantage: A video collection can essentially be treated as an image collection, where each, possibly annotated, image acts as entry point into the video.

4 New Paradigms in Information Visualization

The last decade has witnessed an explosion in interest in the field of information visualization, e.g. [47, 15, 70, 41, 2, 12, 55, 4, 68, 82, 66, 10]. We added three new techniques to the pool of existing visualization paradigms, based on our design studies [5,13]. These techniques all revolve around a representation of documents in the form of bag-of-words vectors, which can be clustered to form groups; we use a variant of the buckshot clustering algorithm for this. Another common element of our visualisations is the notion of *keywords* that are specific to the returned set of documents. The keywords are computed using a simple statistic; for details see [13, 39]. The new methods are:

Sammon Cluster View. This paradigm uses a Sammon map to generate a two dimensional screen location from a many-dimensional vector representing a cluster centroid. This map is computed using an iterative gradient search [63] while attempting to preserve the pairwise distances between the cluster centres. Clusters are thus arranged that their mutual distances are indicative of their relationship. The idea is to create a visual landscape for navigation. Fig 2 shows an example of such an interface. The display has three panels, a scrolling table panel to the left, a graphic panel in the middle and a scrolling text panel to the right that contains the traditional list of returned documents as hotlinks and snippets. In the graphic panel each cluster is represented by a circle and is labelled with its two most frequent keywords. The radius of the circle informs about the cluster size. The distance between any two circles in the graphic panel is an indication of the similarity of their respective clusters: the nearer the clusters, the more likely the documents contained within will be similar. When the mouse passes over the cluster circle a 'tool tip' box in the form of a pop-up menu appears that allows the user to select clusters and *drill down*, ie, re-cluster and re-display only the documents in the selected clusters. The back button undoes this process and