

Progress in Informatics ABSTRACTS

No.6, March 2009

Contents

Special Issue: Leading ICT technologies in the Information Explosion

Guest Editorial

- 1 ——— **Leading ICT technologies in the Information Explosion**
情報爆発時代の先端情報・通信技術
高須 淳宏, 安達 淳

Survey Paper

- 3 ——— **Image retrieval: Research and use in the information explosion**
情報爆発時代における画像の検索と利用の研究動向
井上 雅史

本論文は、情報爆発時代における画像の検索と利用に関する研究における困難と可能性を展望する。大量の画像の存在は、以下の4つの研究項目の重要性を示唆する。1) 画像アクセスと組織化の技術、2) 大規模画像アクセスのための計算基盤、3) 画像利用における人間・計算機間のインタラクション及び人的要素、4) 画像メディアに関わる社会的要素、画像検索システムはかつては実験的な道具であるが、限られた領域で専門家によって使用されるものであった。しかし、大量の画像が存在するようになり、検索に精度や耐規模性が要求されるようになってきた。このために、前二者の研究が必要とされている。さらに、大量の画像の存在は、画像の種類の多様化と利用者の多様化をも伴うため、社会的な文脈に関わる後二者の課題も重要となってきている。

Research Papers

- 15 ——— **Utilization of external knowledge for personal name disambiguation**
外部知識を用いて同姓同名の曖昧解消
Quang Minh VU, 高須 淳宏, 安達 淳

最近ウェブ上に存在するデータが爆発的に増大して、その多量なデータを効率的に処理することが重要な課題となっている。本研究はウェブ上にある人に関する情報に注目し、特に人名検索における同姓同名問題を解決する手法を提案する。人名で検索するときには発生するのが同姓同名問題で、検索結果に複数の人に関連する情報が順不同に混ざって得られる。本手法では同姓同名の問題を解決するために、外部の知識を用いて、人の重要な情報をより効率的に発

見する。具体的には、ウェブディレクトリを外部知識として利用して、latent Dirichlet allocation (LDA) の手法でウェブディレクトリに含まれるトピックを抽出する。そして、抽出されたトピックを用いて検索結果中の人に関する重要な情報を認識して、曖昧性の解消を行う。実際のウェブ文書を用いて実験し、提案手法によりベクトル空間や固有名詞抽出に基づく曖昧性解消手法より良い結果が得られた。

- 27 ——— **Building web page collections efficiently exploiting local surrounding pages**

周辺ページを活用したウェブページコレクションの効率的な構築

Yuxin WANG, 大山 敬三

本論文では局所周辺ページを活用してより少ない人手判定コストで高品質のウェブページコレクションを構築するための方式を示し、例として研究者ホームページを対象カテゴリに用いた実験を通して提案方式の有効性を示している。本方式は簡略フィルタリングと精密分類の二つのプロセスからなっている。両プロセスにおいて、エントリページと周辺ページ間の接続タイプと相対 URL ディレクトリ階層に基づく関係を表現した論理ページグループ構造概念を導入し、この概念に基づき局所周辺ページの内容を利用している。簡略フィルタリングでは、該当する可能性のあるページをウェブから網羅的かつ効率的に収集するため、属性キーワードリストを用いた手法を提案している。ページグループ構造に基づく4種類のページグループモデル (PGM) を用いて周辺ページからキーワードをマージしている。実験結果により、周辺ページを用いない単一ページ手法と比較して、周辺ページをそのまま用いると非該当ページが多く含まれてしまうが、本手法によりその増加を許容可能なレベルに抑えられること、及び単一ページ手法では収集できなかった多くの該当ページを収集できることが示された。精密分類では、精度保証と再現率保証の二つの基本分類器を用いて3段階分類器を構成する手法を提案する。入力は該当、非該当、及び要判定の3つに分類され、要判定とされたページはアプリケーションにより要求されるコレクション品質を確保するため人手判定を行うことになる。各基本分類器はさらに周辺ページ分類器 (SC) とエントリページ分類器 (EC) から構成される。SC では部品ページらしい周辺ページを選択し、EC でエントリページと周辺ページの情報を用いてエントリページを分類する。実験結果より、SC の導入による基本分類器の大幅な性能向上が確認された。また、要判定ページ数の減少度の評価により、提案手法の有効性が確認された。

41 — Academic linkage: A linkage platform for large volumes of academic information

学術情報の統合に向けた大規模リンケージ基盤の構築

相澤 彰子, 高須 淳宏, 深川 大路, 高久 雅生, 安達 淳

本研究では, 学術情報に焦点をあて, 2層構造を持つ情報同定システムを提案する。まず, 同一の事物や人物を参照する断片化した情報をつなぎあわせる情報同定の考え方について述べ, 次に, 特に書誌および研究者の同定機能を組み合わせた情報同定システムを提案する。応用例として, 共著関係ネットワーク分析結果をあわせて示す。

Technical Notes

49 — Computing the potential lexical productivity of head elements in nominal compounds using the textual corpus

テキスト・コーパスを用いた複合名詞主要部の潜在的造語力

影浦 峯

本ノートでは, テキスト・コーパスから複合名詞主要部の潜在的造語力を計算する手法を導入し, 専門分野コーパスを使って複合名詞の主要部として頻出する 10 の要素の潜在的造語力を比較する。これまで複合名詞をめぐる研究は多数なされてきたが, ほとんどが統合的側面を扱ったものであり, 範列的側面を扱ったものはあまりなかった。これに対し, 本ノートでは, テキスト・コーパスから範列的側面の一つである造語力を計算する。これは, テキスト空間から語彙空間上の属性を復元する研究の一環と位置付けられる。

57 — Statistical string similarity model for information linkage

情報リンケージのための統計的字符串類似度

高須 淳宏

本論文は, 情報リンケージに必要な近似マッチングのための統計的な字符串類似度を提案する。このモデルは隠れマルコフモデルに基づいた統計的なモデルで, モデルのパラメータを訓練データから推定することによって, 多様な情報源から得られた情報を適応的に結びつけることを目的としている。本論文では, モデルのパラメータ推定のための EM アルゴリズムに基づいた効率的な学習アルゴリズムを中心に字符串類似度のための統計的なモデルを提案する。

Regular Paper

Research Paper

63 — A distributed information retrieval manner based on the statistic information for ubiquitous services

ユビキタスサービス環境における P2P 型分散検索手法の提案

土屋 健, 澤野 弘明, Marc LIHAN, 吉永 浩和, 小柳 恵一

本研究では Peer-to-Peer (P2P) 型分散環境において, 現在の検索エンジンと同等のコンテンツの意味情報検索を実現するための手法を提案する。ピアの変動が予想される P2P ネットワーク環境では分散検索を実現するためのインデキシング, インデックスの分散配置, そして検索手法といった解決すべき課題に対する三つの手法について述べる。まず, ユーザ参加型オーバーレイネットワーク上のピア間の協調による意味情報を利用したインデキシング手法と, その生成したインデックスの多次元リングモデル論理空間への分散配置手法を明らかにし, 分散配置されたインデックスの検索手法を提案する。提案手法の評価としてシミュレーション及び, エミュレーションを行う。このとき, 提案手法のスケーラビリティと検索の精度の観点からシステムとしての実現性の考察を行い, その結果から小規模から中規模を対象とした柔軟な分散検索システムとしての利用が期待できることを明らかにする。

*Call for Papers***Progress in Informatics No.7*****Special Issue: 3D Image and Video Technology*****Submission Deadline: 31 August 2009**

High-end capturing, processing and rendering devices are currently investigated for various 3D image and video technologies in robotics, visualization, 3DTV, autonomous vehicles, driver assistance, “flying eyes”, intelligent human-machine interfaces, and so forth. In addition, off-the-shelf digital cameras or video recorders allow already to capture high-resolution images and videos at reasonable costs, contributing to the fast introduction of investigated technologies into new products. Computer processing power and storage have achieved a level that real-time processing for 3D images and 3D videos becomes widely possible.

The aim of this special issue is to provide a forum for presenting current research and discussing future research directions in 3D image and video technology for developing academically challenging, and commercially attractive systems in application areas such as mentioned above. This special issue seeks high-quality, original contributions on cutting-edge research, as well as surveys that help to understand research processes and emerging challenges in this area. Topics of interest include basic research in the application areas mentioned above, or in more generic areas such as

- 3D modeling
- 3D image and video analysis
- Object recognition and tracking
- Image-based rendering
- Video classification and clustering
- Video retrieval and browsing
- User-centered visual application for 3D image/video

For more information, contact the **Guest Editors:**

Akihiro Sugimoto

Digital Content and Media Sciences Research Division
National Institute of Informatics, Japan
sugimoto@nii.ac.jp

Yoichi Sato

Institute of Industrial Science
The University of Tokyo, Japan
ysato@iis.u-tokyo.ac.jp

Reinhard Klette

Department of Computer Science
The University of Auckland, New Zealand
r.klette@auckland.ac.nz

* “Progress in Informatics” <http://www.nii.ac.jp/pi/>