A Citation-based approach to automatic topical indexing of scientific literature

Abdulhussain E. Mahdi[1] and Arash Joorabchi

Department of Electronic and Computer Engineering, University of Limerick, Limerick, Republic of Ireland

Appendix A

Note 1. The lists of top five keyphrases assigned to each document by human indexers and Maui are extracted from Olena Medelyan's PhD thesis [1].

Note 2. The number in front of each keyphrase in parentheses represents the number of human indexing teams who have assigned the keyphrase to the document. Also, the keyphrases are sorted in a descending order according to their assignment frequency by human indexing teams or keyphraseness scores by the Maui and our CKE algorithm.

Document ID: 12049

Document title: Occam's razor: the cutting edge for parser technology

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Yacc (13)	Compiler-compiler (9)	occam (6)
Parsing (12)	Yacc (13)	programming languages (4)
Compiler-compiler (9)	Programming language (4)	yacc (13)
Backus Naur form (9)	Parsing (12)	software engineering (0)
Compiler (6)	Compiler (6)	obj3 (0)

[1] *Correspondence to:* Abdulhussain E. Mahdi, Department of Electronic and Computer Engineering, University of Limerick, Limerick, Republic of Ireland. Email: Hussain.Mahdi@ul.ie

Document ID: 10894.

Document title: A safe, efficient regression test selection technique

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Regression testing (15)	Algorithm (7)	regression testing (15)
Software maintenance (13)	Control flow (0)	software maintenance (13)
Control flow graph (10)	Software maintenance (13)	control flow graph (10)
Software testing (9)	Computer software (1)	software testing (9)
Algorithm (7)	Test suite (2)	test suite (2)

Document ID: 13259.

Document title: Cone trees in the UGA graphics system

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Hierarchical model (7)	Visual display unit (0)	information (0)
3D computer graphics (7)	Graphics (0)	visualization (8)
Visualization (graphic) (6)	Computer graphics (3)	xerox parc (0)
Tree (data structure) (5)	Visualization (2)	interactive (0)
Computer graphics (3)	PARC (company) (2)	graphical (0)

Document ID: 16393.

Document title: Cache coherence for shared memory multiprocessors

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Virtual memory (15)	Microprocessor (0)	cache coherence (9)
Multiprocessing (9)	Cache coherence (9)	multiprocessor (9)
Cache coherency (9)	CPU cache (0)	processors (0)
Consistency model (7)	Parallel computing (2)	sequential consistency (6)
Sequential consistency (6)	Sequential consistency (6)	consistency model (7)

Document ID: 18209.

Document title: Mutable object state for object-oriented logic programming

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Object-oriented programming (15)	Object-oriented programming (15)	logic programming (14)
Logic programming (14)	Logic (2)	object oriented programming (15)
Linear logic (6)	Logic programming (14)	linear logic (6)
Immutable object (5)	Prolog (1)	deductive databases (4)
Deductive database (4)	Programming language (1)	prolog (1)

Document ID: 19970.

Document title: A new deterministic parallel sorting algorithm

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Sorting algorithm (13)	Sorting algorithm (13)	algorithm (2)
Parallel computing (10)	Algorithm (2)	parallel computation (10)
Deterministic algorithm (6)	Sampling (statistics) (1)	load balancing (4)
Computational complexity theory (5)	Sampling (music) (1)	parallel algorithms (4)
Load balancing (computing) (4)	Parallel computing (10)	sorting algorithms (13)

Document ID: 23267.

Document title: P++: A language for software system generators?

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Programming language (12)	Programming language (12)	programming language (12)
C++ (6)	Abstraction (2)	software architecture (1)
Component-based software engineering (6)	Abstraction (computer science) (5)	object oriented programming (3)
Encapsulation (6)	Software system (0)	c++ (6)
Abstraction (computer science) (5)	Software engineering (4)	implementation (0)

Document ID: 23507.

Document title: Models for computer generated parody

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Parody (11)	Artificial intelligence (9)	artificial intelligence (9)
Language model (9)	Ernest Hemingway (5)	computational (2)
Artificial intelligence (9)	Computer science (0)	parodies (11)
Vocabulary (7)	Grammar (4)	grammar (4)
Natural language processing (6)	Statistics (3)	computer programming (1)

Document ID: 23596.

Document title: The effect of group size and communication modes in CSCW environments

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Communication (8)	Problem solving (2)	cscw (8)
Computer supported cooperative work (8)	Human communication (1)	computer supported cooperative (0)
Collaborative software (7)	Computer supported cooperative work (8)	human computer interaction (5)
Collaborative workspace (7)	Computer (0)	interaction (2)
Collaboration (5)	Communication (8)	social psychology (0)

Document ID: 25473.

Document title: Extracting multi-dimensional signal features

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Content-based image retrieval (7)	Tree (data structure) (1)	computer vision (5)
Image processing (5)	Wavelet (1)	feature extraction (5)
Computer vision (5)	Discrete cosine transform (2)	color histogram (2)
Image compression (5)	Database (2)	indexing and retrieval (0)
Feature extraction (5)	Color histogram (2)	image compression (5)

Document ID: 287.

Document title: Clustering full text documents

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Machine learning (13)	Machine learning (13)	machine learning (13)
Cluster analysis (10)	Natural language processing (4)	information retrieval (8)
Information retrieval (8)	Natural language (5)	algorithm (2)
Index (search engine) (6)	Algorithm (2)	natural language processing (4)
Natural language (5)	Information retrieval (8)	unsupervised learning (4)

Document ID: 37632.

Document title: The internet software visualization laboratory

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Software visualization (12)	Software visualization (12)	software visualization (12)
Electronic learning (8)	Computer programming (3)	student (1)
Distance education (6)	Visualization (2)	computer program (3)
Education (4)	Computer (0)	internet (4)
Internet (4)	Computer science (0)	eisenstadt (0)

Document ID: 39172.

Document title: Block edit models for approximate string matching

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
NP-complete (12)	Approximate string matching (3)	edit distance (1)
String searching algorithm (10)	Edit distance (1)	algorithm (2)
Computational complexity theory (5)	Algorithm (2)	np complete (12)
Levenshtein distance (5)	NP-complete (12)	molecular biology (3)
String (computer science) (5)	String searching algorithm (10)	dynamic programming (2)

Document ID: 39955.		
Document title: Structured interviews on the object-oriented paradigm		
The 5 Most frequent keyphrases assigned by 15 human indexing teams	**The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)**	**The top 5 keyphrases assigned by CKE algorithm (under condition A)**
Object-oriented programming (15)	Software engineering (6)	software maintenance (5)
Software engineering (6)	Software maintenance (5)	object oriented programming (15)
Structured interview (6)	Object-oriented programming (15)	object oriented paradigm (0)
Interview (6)	Structured interview (6)	software engineering (6)
Software maintenance (5)	Programming language (0)	object oriented (0)

Document ID: 40879.		
Document title: Instance pruning techniques		
The 5 Most frequent keyphrases assigned by 15 human indexing teams	**The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)**	**The top 5 keyphrases assigned by CKE algorithm (under condition A)**
Machine learning (14)	Algorithm (5)	algorithm (5)
Nearest neighbour algorithm (9)	Metric (mathematics) (0)	instance based learning (0)
Algorithm (5)	Machine learning (14)	machine learning (14)
Training set (5)	Learning (0)	classifier (0)
Artificial intelligence (4)	Training set (5)	nearest neighbor rule (0)

Document ID: 7183.		
Document title: The challenge of deep models, inference structures, and abstract tasks		
The 5 Most frequent keyphrases assigned by 15 human indexing teams	**The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)**	**The top 5 keyphrases assigned by CKE algorithm (under condition A)**
Expert system (17)	Abstraction (computer science) (3)	expert systems (17)
Artificial intelligence (12)	Expert system (17)	artificial intelligence (12)
Model (abstract) (6)	Artificial intelligence (12)	knowledge based systems (3)
Abstraction (5)	Scientific modeling (0)	knowledge acquisition (0)
Knowledge base (5)	Medical diagnosis (0)	knowledge base (5)

Document ID: 7502.

Document title: Using introspective reasoning to select learning strategies

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Machine learning (10)	Introspection (7)	case based reasoning (2)
Reasoning (7)	Reasoning (7)	learning (3)
Introspection (7)	Machine learning (10)	knowledge goals (0)
Artificial intelligence (6)	Case-based reasoning (2)	artificial intelligence (6)
Learning (3)	Artificial intelligence (6)	machine learning (10)

Document ID: 9307.

Document title: Specifying and adapting object behavior during system evolution

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Object-oriented programming (13)	Graph (mathematics) (0)	object oriented programs (13)
Software development process (7)	Object-oriented programming (13)	class dictionary graph (0)
Software engineering (7)	Graph theory (2)	object oriented (0)
Class (computer science) (5)	Software engineering (7)	software engineering (7)
Computer-aided software engineering (4)	Graph (0)	object (2)

Document ID: 20782.

Document title: High performance geographic information systems

The 5 Most frequent keyphrases assigned by 15 human indexing teams	The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)	The top 5 keyphrases assigned by CKE algorithm (under condition A)
Geographic information system (15)	Parallel computing (9)	geographic information systems (15)
Parallel computing (9)	Load balancing (0)	algorithm (0)
Distributed Interactive Simulation (9)	Geographic information system (15)	virtual reality (3)
Load balancing (computing) (8)	Algorithm (0)	load balancing (8)
Parallel programming model (4)	Message passing (1)	parallel computing (9)

Document ID: 43032.		
Document title: Observations and recommendations on software internationalization		
The 5 Most frequent keyphrases assigned by 15 human indexing teams	**The top 5 keyphrases assigned by Maui (Bagged Decision Trees & best features)**	**The top 5 keyphrases assigned by CKE algorithm (under condition A)**
Internationalization and localization (15)	Character encoding (4)	x open portability guide (0)
User interface (6)	User interface (6)	user interface (6)
Software engineering (6)	Internationalization and localization (15)	user interface design (1)
Translation (5)	Interface (computer science) (1)	software (0)
Linguistics (5)	Programming language (0)	users (0)

References

[1] Medelyan, O. *Human-competitive automatic topic indexing*. The University of Waikato, 2009. [Online]. Available: http://adt.waikato.ac.nz/public/adt-uow20091029.160923