JOURNAL OF LAW AND INFORMATION SCIENCE

Published by the Faculties of Law, and Mathematical and Computing Sciences
New South Wales Institute of Technology

Vol. 1 No. 2 **1982**

EDITORIAL BOARD

Chairman

Hon. Mr. Justice M.D. Kirby,
Chairman, Australian Law Reform Commission

Editor

Dr. R.A. Brown,
Lecturer in Law, N.S.W.I.T.

Members

Mr. G.W. Bartholomew,
Dean, Faculty of Law,
N.S.W.I.T.

Professor J. Bing,
Norwegian Research Centre for
Computers and Law, Oslo University

Professor P. Catala,
Institut de Recherches et d'Etudes
pour le Traitement de l'Information
Juridique, Université de Montpellier I

Professor J. Goulet,
Faculté de Droit
Université Laval, Quebec

Mr. W.A. Steiner,
Librarian
Institute of Advanced Legal Studies

Mr. R.J. Watt,
Senior Lecturer in Law,
N.S.W.I.T.

Mr. D. Biles,
Assistant Director (Research)
Australian Institute of Criminology

Professor A. Blackshield,
Dept. of Legal Studies
La Trobe University

Dr. V.X. Gledhill,
Dean, Faculty of Mathematical
and Computing Sciences,
N.S.W.I.T.

Mr. C.H. Gray,
Dept. of Mathematics and Statistics,
C.S.I.R.O.

Mr. C. Tapper,
All Souls Reader in Law,
Magdalen College, Oxford

Professor D.N. Weisstub,
Professor of Law and Psychiatry,
Osgoode Hall Law School

Professor D. Whalan,
Faculty of Law,
Australian National University

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

This volume may be cited as
(1982) 1 J.L.I.S.

Articles, books for review, subscriptions and all inquiries should be addressed to The Editor, Journal of Law and Information Science, C/O Faculty of Law, New South Wales Institute of Technology, P.O. Box 123, Broadway, N.S.W. 2007, Australia.

Copyright © 1982 New South Wales Institute of Technology

All rights reserved. Subject to the law of Copyright no part of this publication may be reproduced stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise, without the permission of the owner of the copyright. All inquiries seeking permission to reproduce any part of this publication should be addressed in the first instance to The Editor, Journal of Law and Information Science, c/o Faculty of Law, New South Wales Institute of Technology, P.O. Box 123, Broadway, N.S.W. 2007, Australia.

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

CONTENTS

		Page
Editorial		vii
Articles		
Legality — Information Technology and the Laws of Evidence	*T.H. Smith*	89
Theories of Information in Law (Recent Developments in the Discipline of Rechts- und Verwaltungsinformatik -RVI- in Germany) A Background Paper	*H. Burkert*	120
The Use of Citation Vectors for Legal Information Retrieval	*C. Tapper*	131
Computerisation of Legal Material in Australia	*P.J. Ward*	162
Casenote		
Conwell v. *Tapfield*	*D.J. Robinson*	175
Book Reviews		
The Solicitor and the Silicon Chip	*R.A. Brown*	179
Control & Audit of Small/Medium Computer Systems	*R.J. Watt*	180

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

THE USE OF CITATION VECTORS FOR LEGAL INFORMATION RETRIEVAL

C. TAPPER

Colin Tapper is one of the founders of the study of computers and law in England, and this paper adds to his contributions to the field. As its title indicates, the article is concerned with the use of case citations as selection vectors in legal information retrieval, and, in particular, with the value of citation vestors in comparison to the usual semantic vectors currently used.

The author details recent experiments with citation vectors in the United States and at the Norwegian Research Centre for Computers and Law (NRCCL). The comparative results of using citation vectors against semantic vectors in these experiments are documented and considered, and Mr. Tapper provides some valuable discussion of the algorithms used in computing and assessing vectors in data retrieval. Despite the complexity of this work, it will be of great value to all interested in the field of computers and law, because of its implications for the future development of legal data retrieval.

The first section of this article is intended for those who have no, or little, previous awareness of legal information retrieval techniques. Since the main aim of the article is to explain the theory behind the substitution for such methods of citation vectors those who have the requisite familiarity with matching and vector based systems as applied to law might prefer to start with the second section.

1. *Current Legal Information Retrieval Techniques*

It is now about 25 years since at the University of Pittsburgh in Pennsylvania Professor John Horty first succeeded in applying computerised methods to the retrieval of legal information. It is a tribute to his insight that the techniques which he devised remain the bedrock of virtually all of the systems which operate in the world today. The essence of the technique is the identification in the text of a document of a word, or words, in a particular combination which have been selected by the lawyer as being likely to indicate the relevance of that document to the lawyer's problem. As normally implemented the system creates a concordance of the full legal texts constituting the database of the system, excluding only words of such low *prima facie* information content that they are highly unlikely to be nominated by lawyers as search terms. Each concordance item then becomes a potential search term, and searches are typically conducted by the nomination of classes of words, for example synonyms, grammatical variations, particularisations and generalisations, which must occur in a given relationship to other similar classes in a docu-

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

relevant document. So as to accomplish this process the lawyer must first accustom himself to thinking in terms of word occurrence rather than directly in terms of the meaning of a document. He must be comprehensive in his classification of terms, and he must be able to specify the appropriate logical relationship in terms of Boolean logic and relative sequential occurrence in order to secure an answer. In a commercially operational system he will also be well-advised to consider very carefully not only whether his categorisation is appropriate, but whether it is the most efficiently appropriate formulation of his search, since the more efficient the search the quicker and cheaper it becomes.

It is no exaggeration to say that this process teems with problems both for the system designer and for the average lawyer. Many of these can be mitigated by proper training and continual practice. Some of them are more intractable. At the level of the selection of words with *prima facie* low information content there is the difficulty that "word" is strictly speaking an inaccurate designation. "Words" in the system also encompasses such things as numbers and abbreviations, and would more properly be described as strings of characters. In this extended sense it is rarely possible to predict with certainty that a given string has no information content. Most systems for example exclude the string "A" on the basis that the upper case indefinite article is rarely essential to a search. This may be true, but it is not sufficient to justify the exclusion of the string "A" from the concordance since "A" does have meaning in some contexts, for example, the Australian abbreviation "A" followed immediately by "L" followed immediately by "R". In the United States "A" is itself an abbreviation for an important series of reports. It is, of course, immaterial that the abbreviation occurs only in some other jurisdiction if material from that jurisdiction can ever be reported in one's own.

Semantics and syntax present further difficulties. A basic problem of a semantic nature is that character strings may not denote concepts uniquely or exclusively. In many contexts the strings "minor", "infant", "child", "juvenile", "boy" and "girl" are equivalent, in others they are not. Conversely strings like "office", "bank", "safe", "deposit" and "flag" have more than one meaning. In the former case one of the problems is to think of all of the possible alternatives so as to include them in the search formulation, in the latter it is to think of them so as to draft the combination of classes in such a way as to exclude the unintended meanings. Given the presumptively inclusive range of search terms this can be extremely difficult, thus in one search of British material when seeking documents relevant to the Gas Board it was found that the string "Gas" was ambiguous because there had been an Indian litigant in one case of that name. To some extent these problems interact with each other, for example when in an effort to avoid the former problem so far as grammatical variants are concerned truncation is used, that is, specifying a word root followed by a special character to retrieve all strings commencing with that root, extra problems are created in relation to unanticipated

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.