

Fusion and Confusion in Electronic Discovery for the Law

Dave Lewis

David D. Lewis Consulting

Talk presented at *Fusion Fest: Marrying Methods to Solve Real-world Problems*, DIMACS,
Rutgers University, New Brunswick, NJ, October 11, 2014.

Simulated Entity Resolution by Diverse Means: DIMACS Work on the KDD Challenge of 2005

by

Andrei Anghelescu, Aynur Dayanik, Dmitriy Fradkin,
Dept. of Computer Science
Rutgers University
New Brunswick, New Jersey 08854

Alex Genkin, DIMACS Rutgers University New Brunswick, New Jersey 08854	Paul Kantor, SCILS Rutgers University New Brunswick, New Jersey 08854
---	--

David Lewis, David D. Lewis Consulting	David Madigan, Department of Statistics Rutgers University New Brunswick, New Jersey 08854
---	---

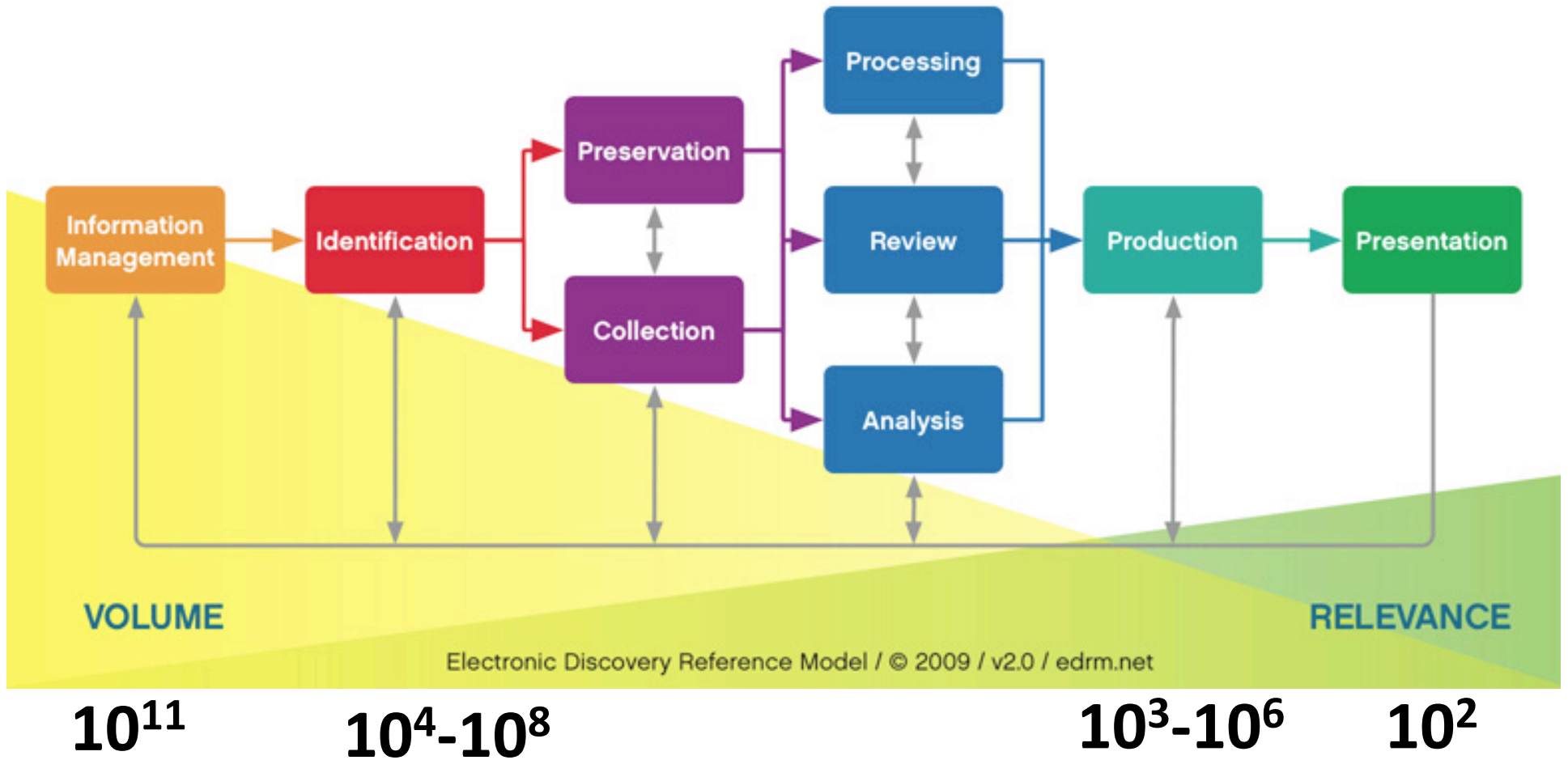
Ilya Muchnik, Fred Roberts,
DIMACS
Rutgers University
New Brunswick, New Jersey 08854

Discovery

discovery n. the entire efforts of a party to a lawsuit and his/her/its attorneys to obtain information before trial through ***demands for production of documents,...***

law.com

Electronic Discovery Reference Model



Where the Money Goes

Understanding Litigant Expenditures for
Producing Electronic Discovery

Nicholas M. Pace, Laura Zakaras



RAND

Institute for Civil Justice

A RAND LAW, BUSINESS, AND REGULATION INSTITUTE

for maintaining a steady income stream.⁸ Arguably one of the least glamorous segments of the legal services industry, the use of large numbers of temporary licensed attorneys to perform eyes-on review has blossomed in recent years, with what appears to be a corresponding increase in complaints about pay and working conditions voiced by practitioners who take such positions.⁹ In New York City, a location known for relatively high costs of living, licensed attorneys working on a contract basis in a large-scale review operation might receive no more than \$25 per hour without benefits, be expected to review 80 to 100 documents per hour, have no expectations that the position will last more than a few weeks or months (early terminations of review projects due to settlement are frequent occurrences), and are sometimes required to work in excess of 12 hours per day without overtime pay.¹⁰ Although the apparent oversupply of attorneys is likely to last for some time, it is not realistic to expect that hourly rates will continue to plunge indefinitely.¹¹ Even if high-volume document-review projects migrate to areas

The New York Times

Front page, Mar.
4, 2011

SMARTER THAN YOU THINK

Armies of Expensive Lawyers, Replaced by Cheaper Software



Ramin Rahimian for The New York Times

"People get bored, people get headaches. Computers don't," said Bill Herr, a lawyer who used to work for a chemical company.

By [JOHN MARKOFF](#)

Published: March 4, 2011

When five television studios became entangled in a Justice Department antitrust lawsuit against [CBS](#), the cost was immense. As part of the obscure task of "discovery" — providing documents relevant to a lawsuit — the studios examined six million documents at a cost of more than \$2.2 million, much of it to pay for a platoon of lawyers and paralegals who worked for months at high hourly rates.

[f](#) RECOMMEND

[t](#) TWITTER

[c](#) COMMENTS
(245)

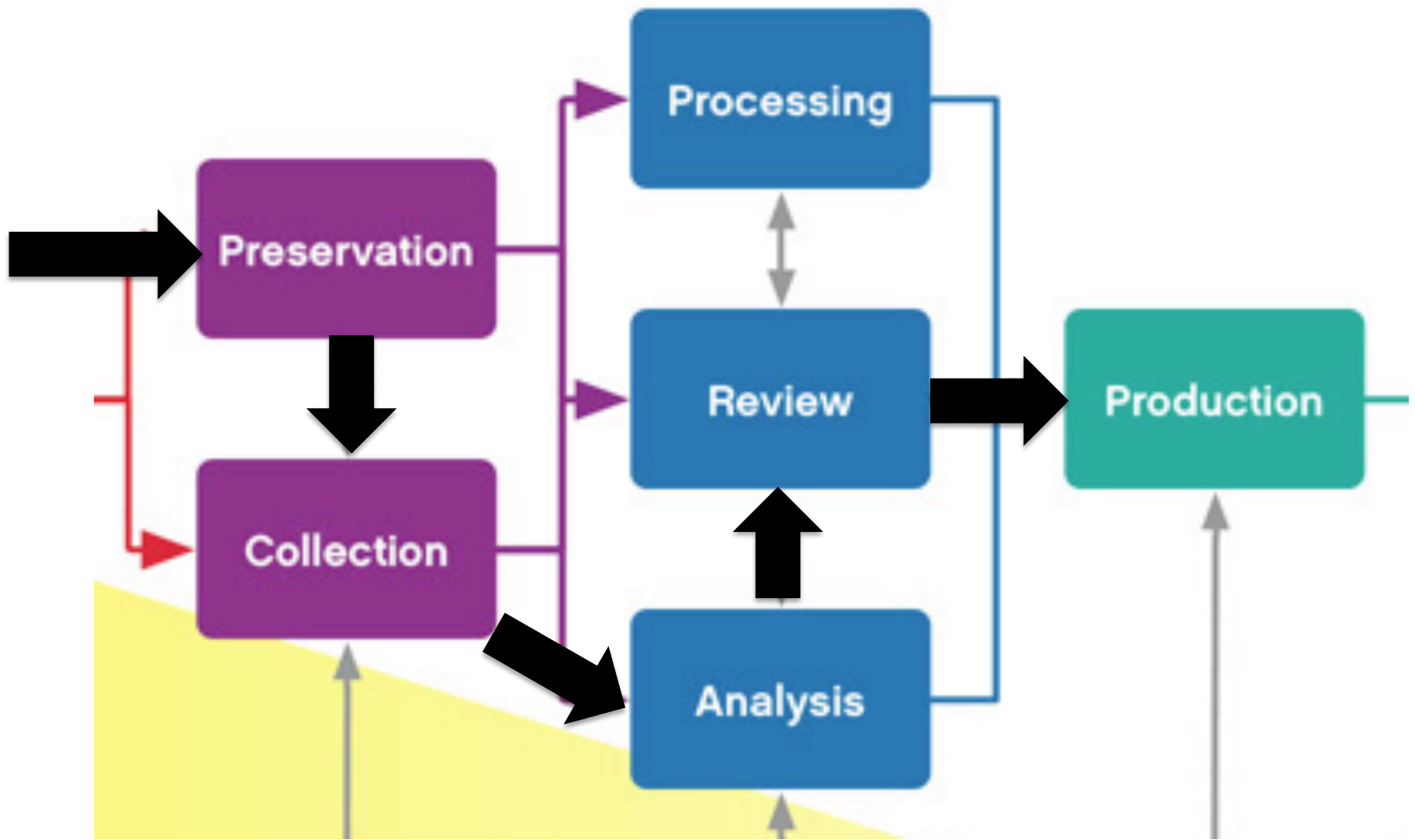
[e](#) E-MAIL

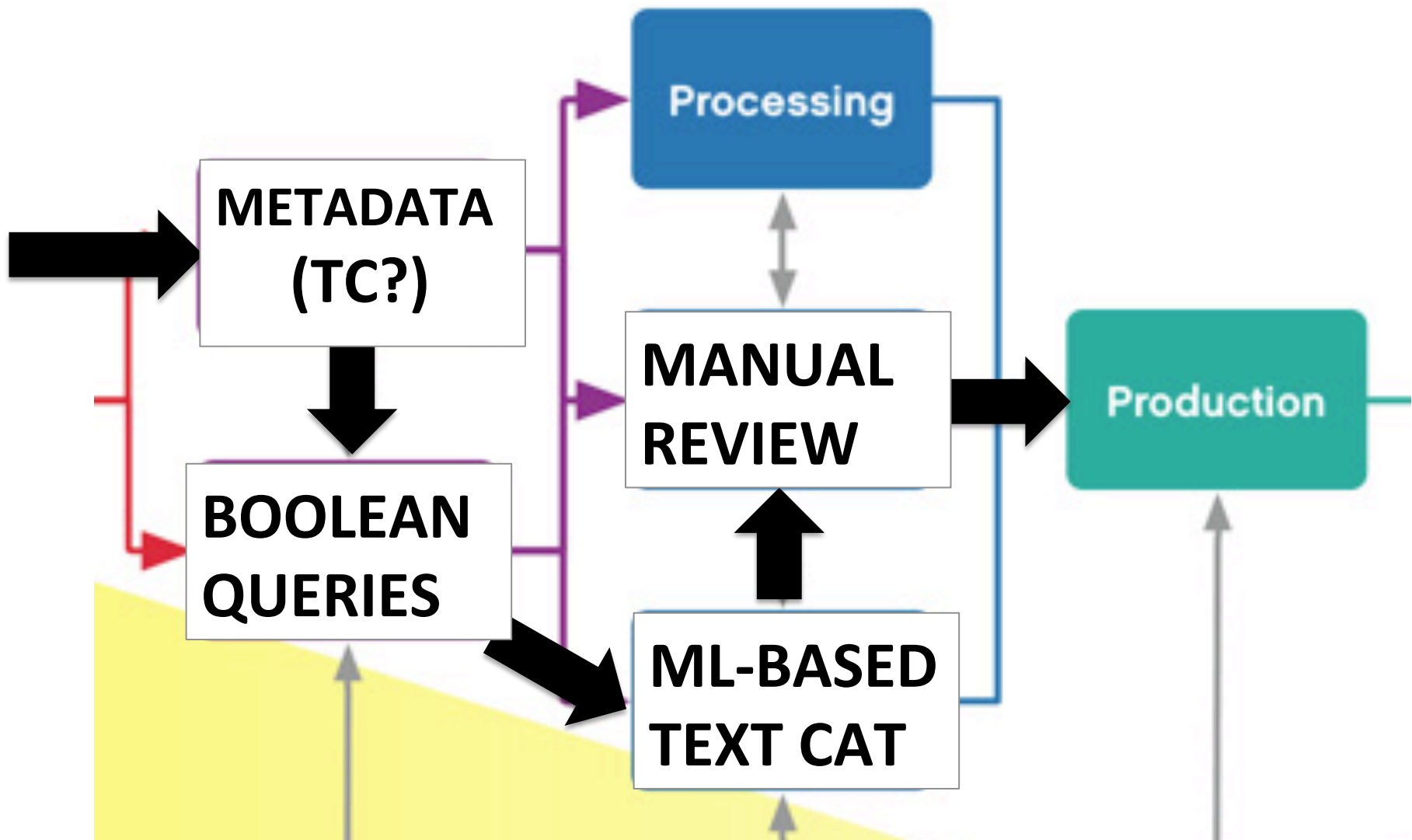
[p](#) PRINT

[s](#) SINGLE PAGE

[r](#) REPRINTS

[+](#) SHARE





Controversies

- Whether / how to evaluate
 - **Classifiers produced by machine learning**

Controversies

- Whether / how to evaluate
 - Preservation
 - Collection
 - Culling by Boolean keyword queries
 - **Classifiers produced by machine learning**
 - Manual review by teams of lawyers
- Battle fought in technical literature, law review articles, court debates, rulemaking,...

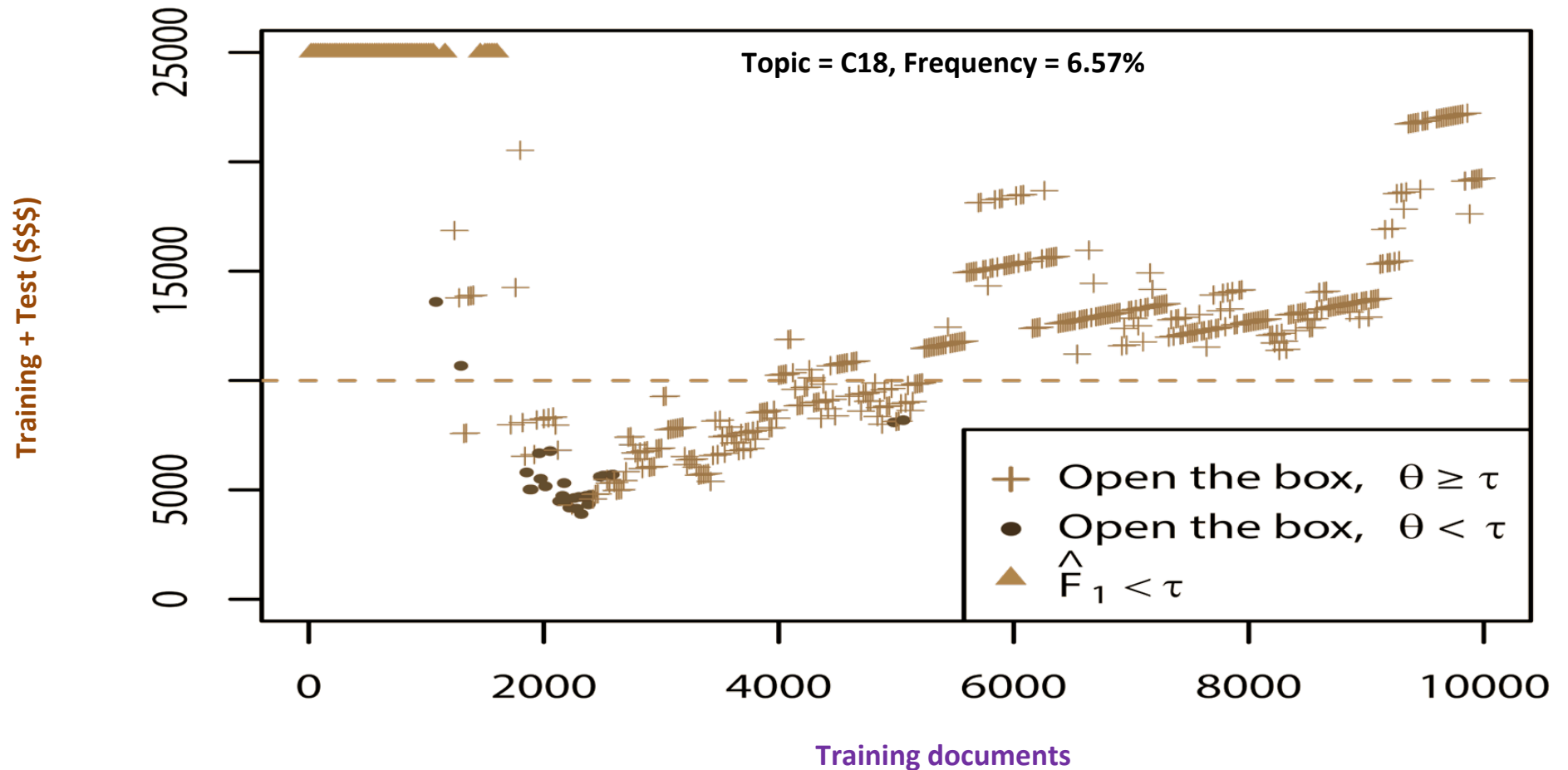
“Proportionality” (Cost-Benefit Analysis)

(C) *When Required.* On motion or on its own, the court must limit the frequency or extent of discovery otherwise allowed by these rules or by local rule if it determines that:

(iii) the burden or expense of the proposed discovery outweighs its likely benefit, considering the needs of the case, the amount in controversy, the parties' resources, the importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.

FEDERAL RULES
OF
CIVIL PROCEDURE

Minimizing Combined Cost of Training, Testing, and Manual Review



Joint work with Oard, Webber, Bagdouri (SIGIR 2013, CIKM 2013)

Studying the Cost and Value of Library and Information Services: Applying Functional Cost Analysis to the Library in Transition

Eileen G. Abels

College of Library and Information Services, 4105 Hornbake Library Building, South Wing, University of Maryland, College Park, MD 20742. E-mail: ea29@umail.umd.edu

Paul B. Kantor and Tefko Saracevic

*Alexandria Project Laboratory, School of Communication, Information, and Library Studies, Rutgers, the State University of New Jersey, New Brunswick, NJ 08903.
E-mail: kantor@zodiac.rutgers.edu; saracevic@zodiac.rutgers.edu*

Brief Communication

A Model for the Stopping Behavior of Users of Online Systems*

Paul B. Kantor

Tantalus, Inc., 3257 Ormond Road, Cleveland, Ohio 44118

MANAGEMENT SCIENCE

Vol. 37, No. 3, March 1991

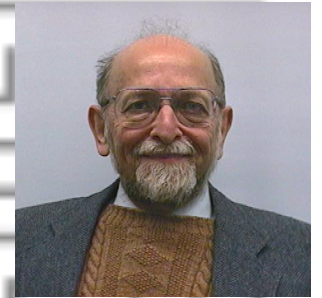
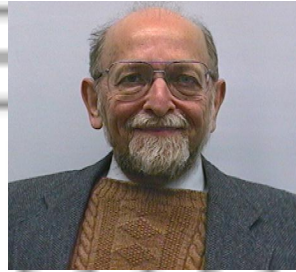
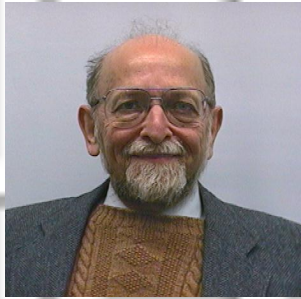
Printed in U.S.A.

THEORETICAL FOUNDATION FOR A LEARNING RATE BUDGET*

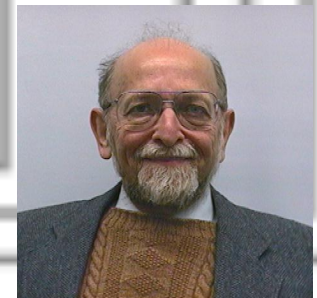
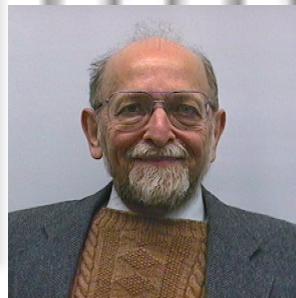
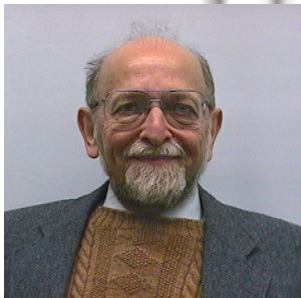
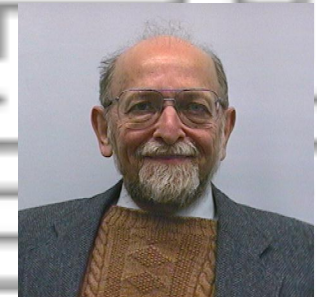
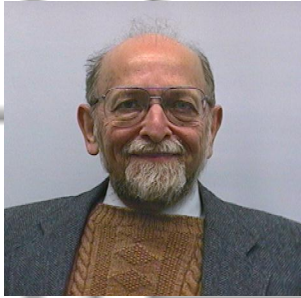
PAUL B. KANTOR AND WILLARD I. ZANGWILL

Tantalus, Inc., 3257 Omond Road, Cleveland, Ohio 44118

Graduate School of Business, University of Chicago, Chicago, Illinois 60637



YOU ARE IN A
MAZE OF TWISTY
LITTLE PASSAGES,
ALL ALIKE



Back to the Future

- The law is a new frontier for information system design...
- ...but we already have the start of many trails blazed by Paul Kantor and his colleagues