How Accurate is Item-Level Search for Digital Records?

Douglas W. Oard and Jason R. Baron

UNIVERSITY OF MARYLAND 18 56 — COLLEGE OF INFORMATION STUDIES

Digital Records

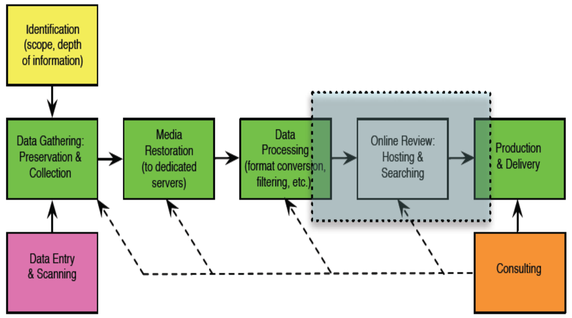
- Rapidly increasing scale
- Decentralized creation and management
- Somewhat "self-describing"
  - Content of individual items
  - Relationships between items

Key Questions

- How many "relevant" items exist?
- How many can practically be found?
- Which "methods" work best?
  - Which automated techniques?
  - Which ways of using those techniques?

Applications

- eDiscovery
- FOIA
- End-user search

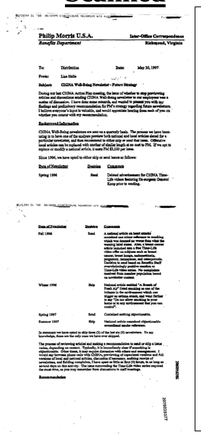


Jack G. Conrad, "E-Discovery Revisited: A Broader Perspective for Researchers," DESI-1

The TREC Legal Track (annually since 2006)

~7 Million Scanned Business Records (Tobacco MSA)

Scanned



OCR

Philip Moxx's. U.S.A. x.dr~am~c.
cvrrespoaa.aa
Benffrts Departmext Rieh>pwna, Yfe&ia
Ta: Dishlbutfon Data aday 90,1997.
From: Lisa Fislla
Sabj.csr CIGNA WeWedng Newsbttsr -
Yntsre StratsU
During our last CIGNA Aatfoa Plan
meadng, tlu iasuo of wLetSae to i0op
per'Irw+ng
artieles aod discontinue mndia6 CIGNA
Well-Being aawslener to om employees was
a
msiter of disanision . I Imvm done somme
reaearc>>, and wanted to pruedt you with
my
Sadings and pcdiminary recwmmeadatioa
for PM's atratezy Ieprding l4aas aewelattee*
.
I believe .vayone'a input is valusble, and
would epproolate hoarlng fmaa aaeh of you
on
whetlne you concur with my
reeommendatioa
…

Metadata

Title: *CIGNA WELL-BEING NEWSLETTER - FUTURE STRATEGY*

Organization Authors: *PMUSA, PHILIP MORRIS USA*

Person Authors: *HALLE, L*

Document Date: *19970530*

Document Type: *MEMO, MEMORANDUM*

Bates Number: *2078039376/9377*

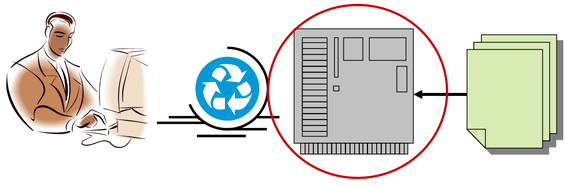
Page Count: *2*

Collection: *Philip Morris*

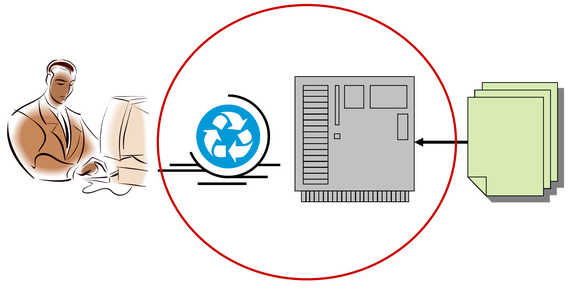
~0.5 Million Corporate Emails w/attachments (Enron)

```
Enron Corp. CEO: Core Businesses Are In 'Great Shape'
Dow Jones News Service, 03/23/01

Enron Says $3.1 Bln Portland General Sale Probably Won't Close
Bloomberg, 03/23/01



USA: Broadband unit in strong shape, says Enron CEO.

03/23/2001
Reuters English News Service
(C) Reuters Limited 2001.

HOUSTON, March 23 (Reuters) - Enron Corp.'s broadband unit will execute
nearly twice as many trades in the first quarter of 2001 as it did in the
fourth quarter of 2000, Enron President and Chief Executive Officer Jeff
Skilling said on Friday.
"My expectation is that we will see over 400, maybe over 500 transactions
this quarter which shows this business is absolutely developing, it is ahead
of plans," Skilling said in a conference call that sought to assuage investor
fears about Enron's stock and its broadband unit specifically.
That is more than all 321 transactions done last year, and about twice the
236 done in the last quarter of 2000, he said.
Skilling also said long supply and a weak telecom market is offering Enron
```

Two Types of Research Designs



"Automatic" Task:
Boolean, Ad Hoc, Relevance Feedback, Batch, Learning

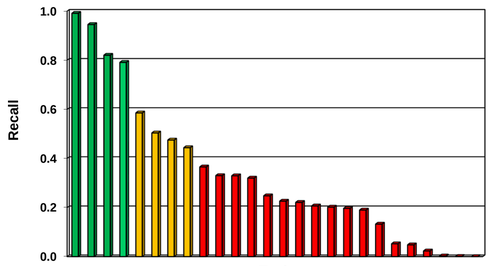


"Interactive" Task

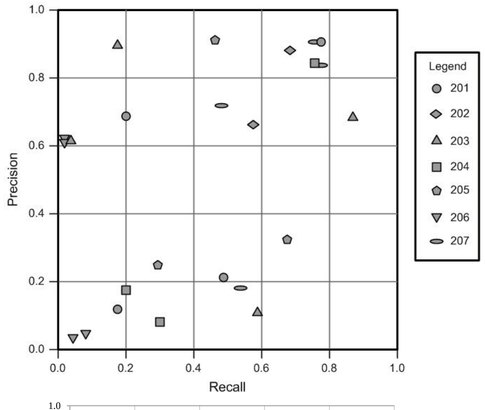
Answer Keys for 116 eDiscovery "Production Requests"

All documents or communications that describe, discuss, refer to, report on, or relate to whether the Company had met, or could, would, or might meet its financial forecasts, models, projections, or plans at any time after January 1, 1999.

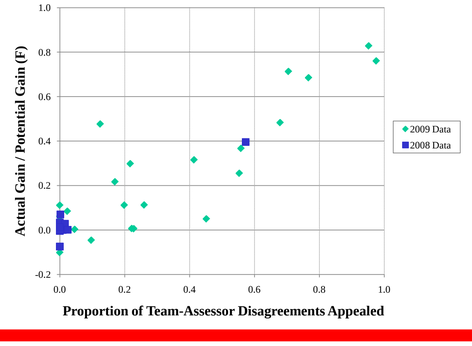
Research Results from 21Companies and Universities in 7 Countries



Negotiated Boolean



Post-Appeal Interactive



Effect of Appeals