

Technology Assisted Review (TAR)

Data Science in Context

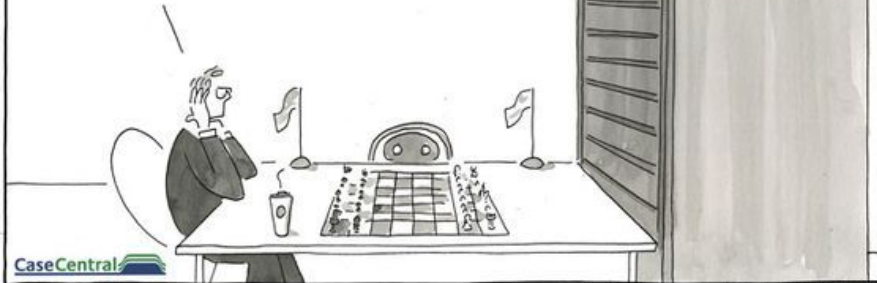
CASE IN POINT

by Tom Fishburne

MAN VERSUS MACHINE

NOW THAT I HAVE COFFEE,
I'M READY TO CHALLENGE
YOU TO REVIEW 10
TERABYTES OF EMAIL

CHECKMATE
/



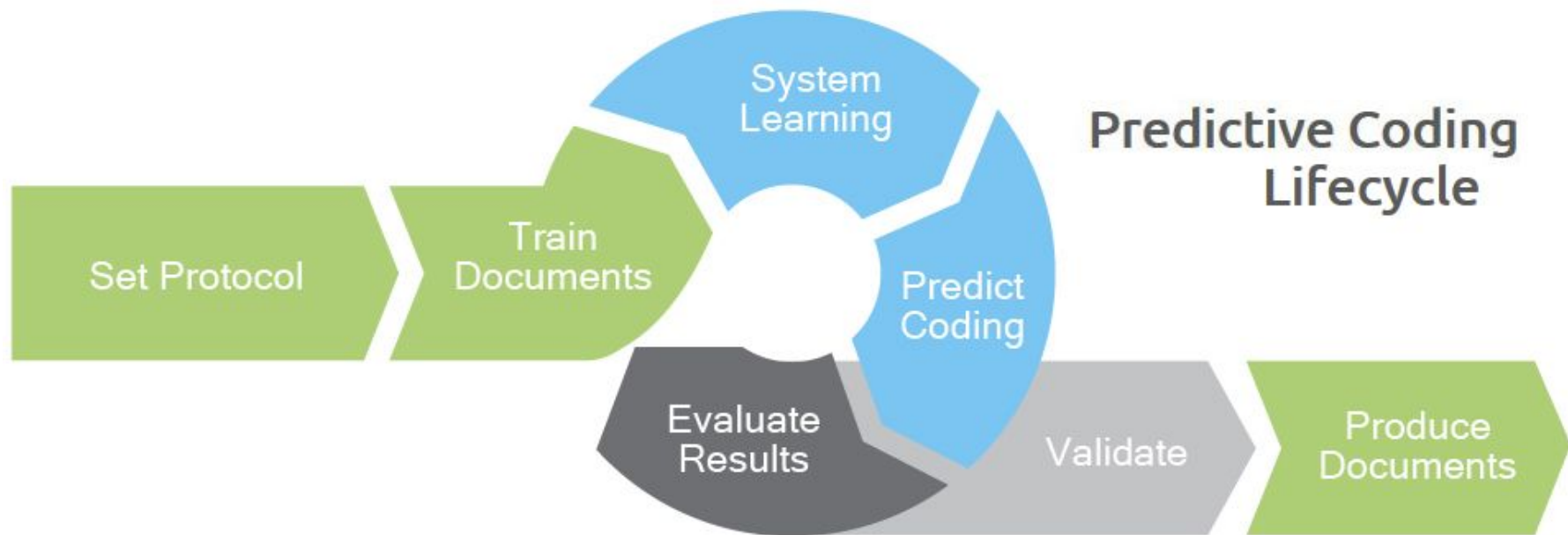
CaseCentral

© 2012

CASECENTRAL.COM / CASE IN POINT

Definitions

- **Richness** - percentage of relevant documents in a document set
- **Recall** - ratio of identified relevant documents to total number of relevant documents
- **Precision** - ratio of identified relevant documents to total number of documents reviewed
- **F-measure** - weighted average of precision and recall



Workflow (Simple Active Learning)

- **Assessment Phase** - Subject Matter Expert (SME) reviews about 500 randomly selected documents
- **Training Phase** - SME reviews roughly 40-50 sets of 40 documents each, which are selected from the ones the algorithm is least certain about
- **Batch Calculation** - System generates relevance score for all documents
- **Testing Results** - Team validates machine learning model

DEMO

equivio
zoom in. find out

<http://www.equivio.com>

Key Points for System Training

- Attorney training the system must be a subject matter expert
- Ideally, only one attorney is training the system
- Training relies purely on document text
- Following should not be considered in relevance calls:
 - Dates
 - Names
 - Numbers
 - Formatting

Benefits

- **Cheaper** - Assuming review rate of 60 docs per hour and \$80 per review hour, eliminating 100,000 documents from the review results in savings of over \$130,000
- **Quicker** - Similarly, eliminating 100,000 documents from the review, cuts down review by over 1,660 hours (or over 16 days for a team of 10 reviewers working 10 hours per day)
- **More Accurate/Consistent** - Software identified over 95% of relevant documents while human reviewers identified on average only 51.1% of relevant documents

Technical Drawbacks

- Cannot handle non-textual files - video, audio, photographs, etc.
- Problem assigning scores to spreadsheets
- Garbage in, garbage out
- Problem identifying privileged material

Legal Challenges

- **Defensibility** - Not yet a standard practice; some teams are hesitant to use it due to uncertainty about court's or opposing counsel's reaction; limited case law
- **Transparency** - Varying degree of transparency, but opposing counsel may seek access to training/seed set
- **Technical Unfamiliarity** - “Four years from now, I predict lawyers who are hesitant to use TAR will be using it—if only because there will be newer, better technology they should be using.” *Judge Andrew Peck, Feb. 2016*

Resources

Evaluation of Machine-Learning Protocols of Technology-Assisted Review in Electronic Discovery

Gordon V. Cormack, Maura R. Grossman. July 2014

<http://www.wlrk.com/webdocs/wlrknew/AttorneyPubs/WLRK.23339.14.pdf>

e-Discovery Team Blog

Ralph Losey

<https://e-discoveryteam.com/>

Predictive Coding a.k.a. Computer Assisted Review a.k.a. Technology Assisted Review (TAR) - Da Silva Moore Revisited

Judge Andrew J. Peck's Opinion and Order, *Rio Tinto PLC v. Vale S.A., et al.* (14 Civ. 3042)

<http://ctrlinitiative.com/114cv3042/>

Thank You
