

The logo consists of the letters 'H5' in a white, sans-serif font, centered within a dark blue square. This square is positioned in the middle of a horizontal line that tapers at both ends, extending across the width of the slide.

H5

# **Finding a Better Way to Search: Benchmarking E-Discovery Methods**

H5 Webinar

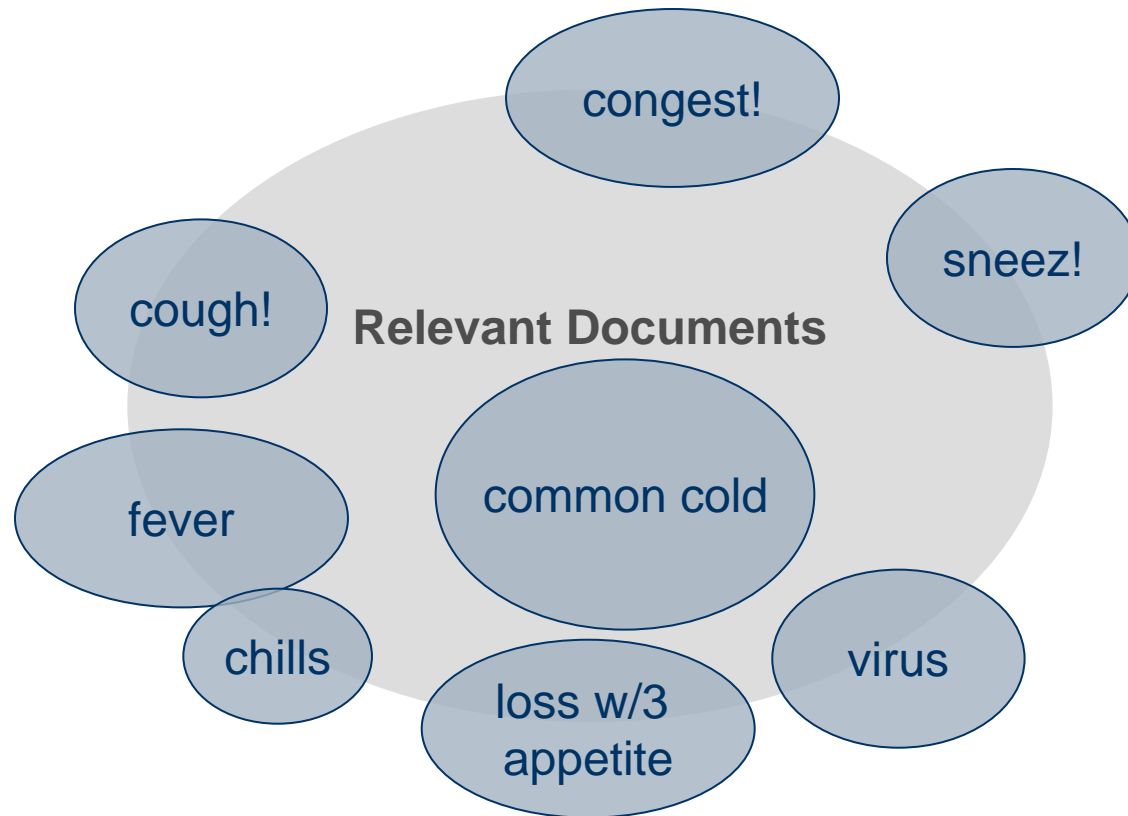
December 10, 2008

- **Moderator:**
  - **Robert Bauer, Ph.D.**, Chief Technology Officer, H5
- **Panelists:**
  - **Jason R. Baron**, Director of Litigation, National Archives and Records Administration; Founding Co-Coordinator, TREC Legal Track
  - **Julia Brickell**, Executive Managing Director and General Counsel, H5
  - **Judge Paul W. Grimm**, Chief Magistrate Judge for the United States District Court for the District of Maryland
  - **Maura R. Grossman**, Counsel, Wachtell, Lipton, Rosen & Katz

- Today's search and review challenges in e-discovery
- Overview of search and review methods and technologies
- Evolving perspectives from the bench
- Overview of benchmarking initiatives (TREC)
- Q&A

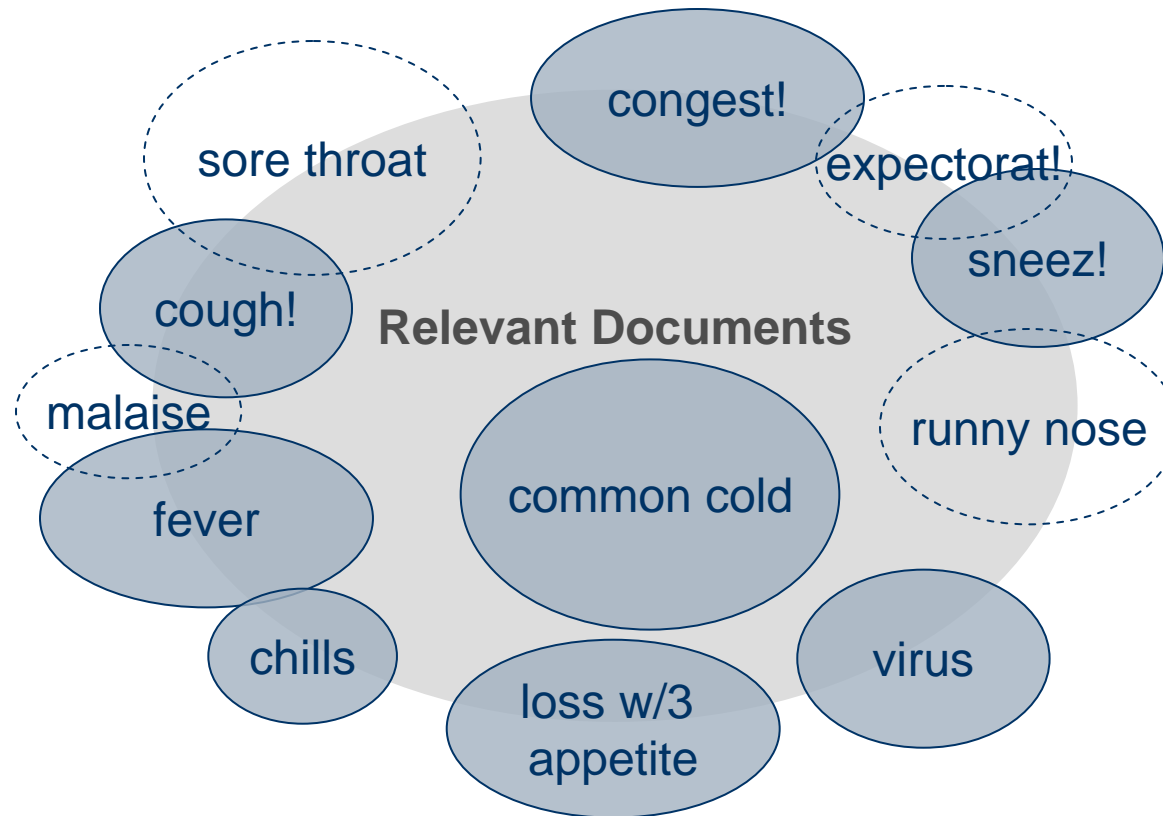
- Today's search and review challenges in e-discovery
  - Overview of search and review methods and technologies
  - Evolving perspectives from the bench
  - Overview of benchmarking initiatives (TREC)
  - Q&A

- ✓ Today's search and review challenges in e-discovery
- **Overview of search and review methods and technologies**
  - Evolving perspectives from the bench
  - Overview of benchmarking initiatives (TREC)
  - Q&A



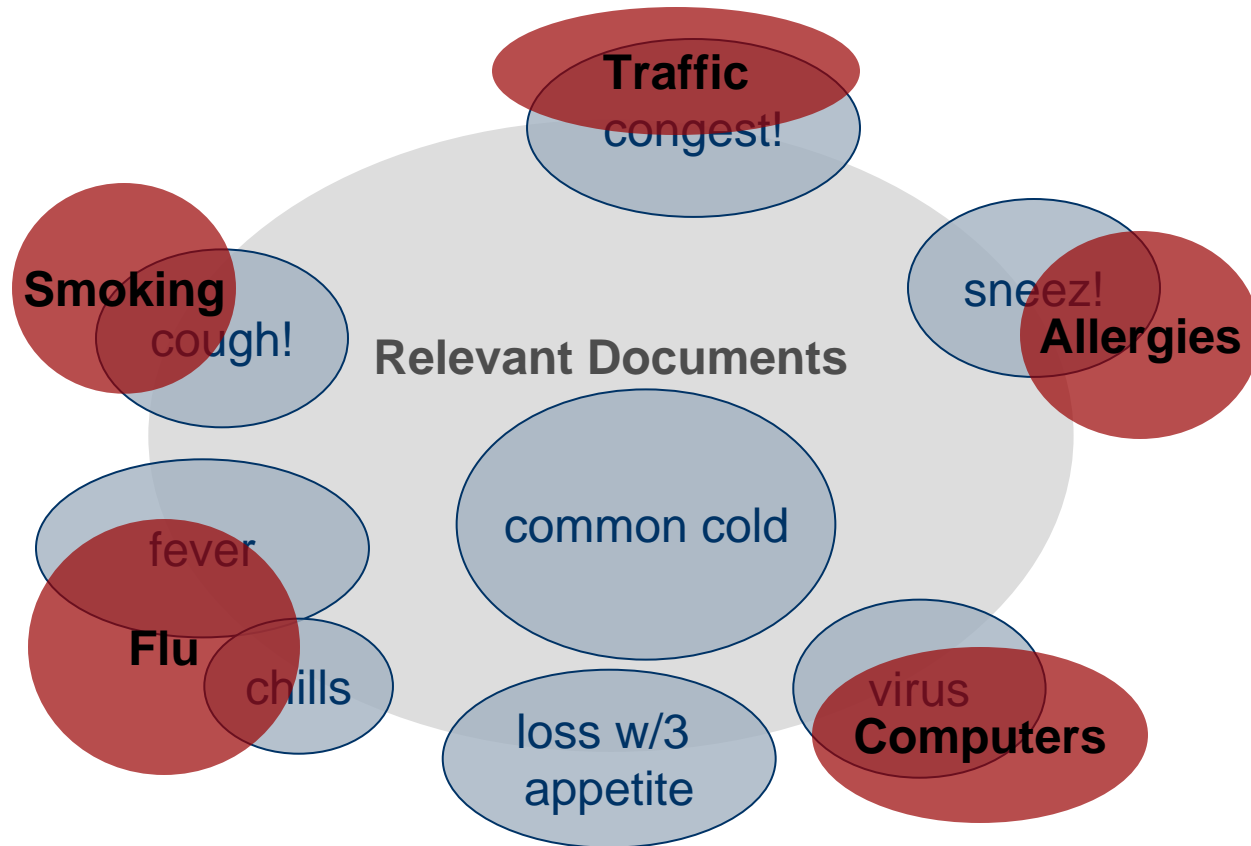
[http://www.theseconconference.org/content/miscFiles/Best\\_Practices\\_Retrieval\\_Methods\\_revised\\_cover\\_and\\_preface.pdf](http://www.theseconconference.org/content/miscFiles/Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf)

# Keyword/Boolean Search: Under-inclusive



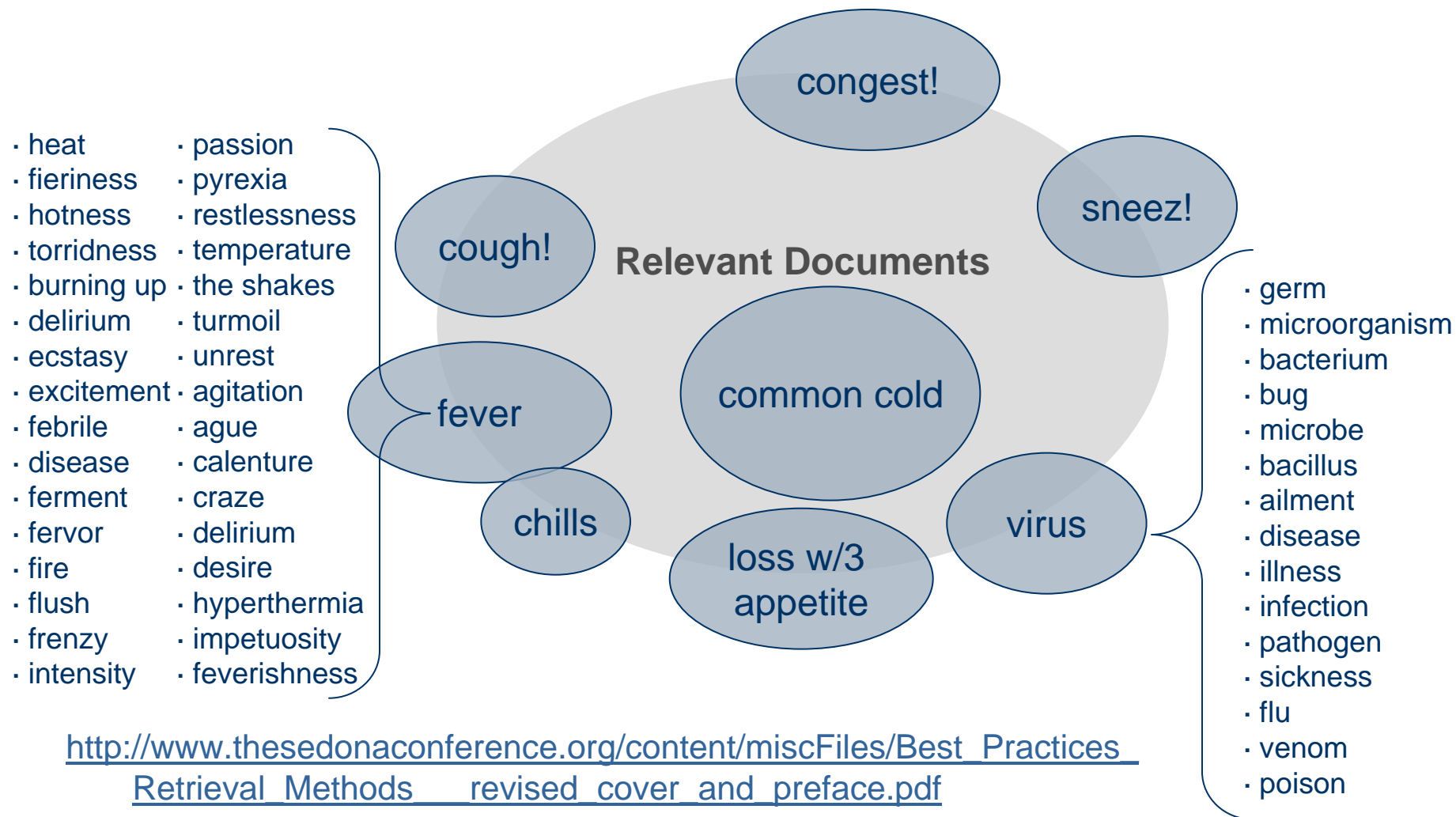
[http://www.theseconferencetalks.org/content/miscFiles/Best Practices Retrieval Methods revised cover and preface.pdf](http://www.theseconferencetalks.org/content/miscFiles/Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf)

# Keyword/Boolean Search: Over-inclusive

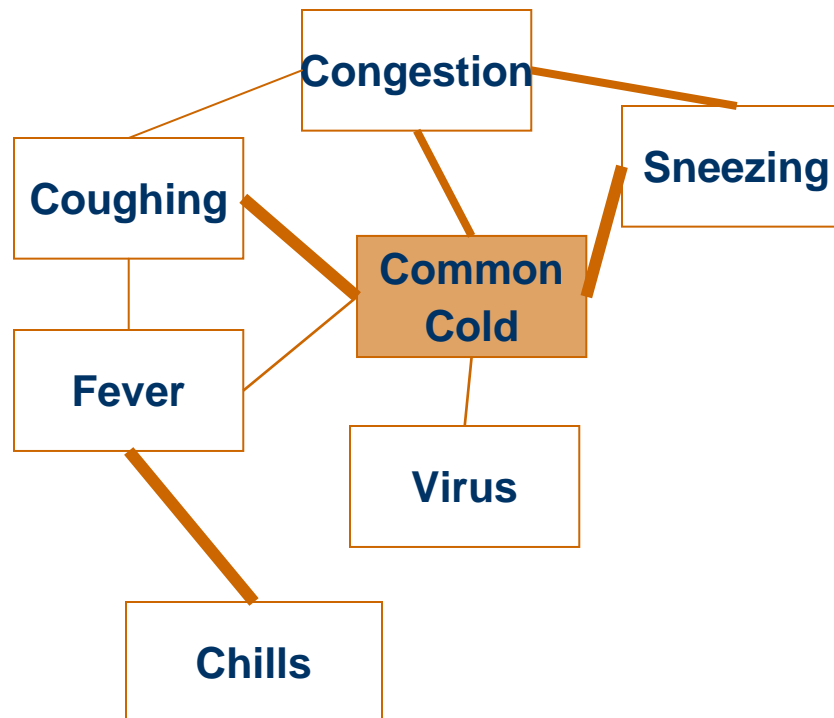


[http://www.theseconconference.org/content/miscFiles/Best\\_Practices\\_Retrieval\\_Methods\\_revised\\_cover\\_and\\_preface.pdf](http://www.theseconconference.org/content/miscFiles/Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf)

# “Concept” Search: Thesauri

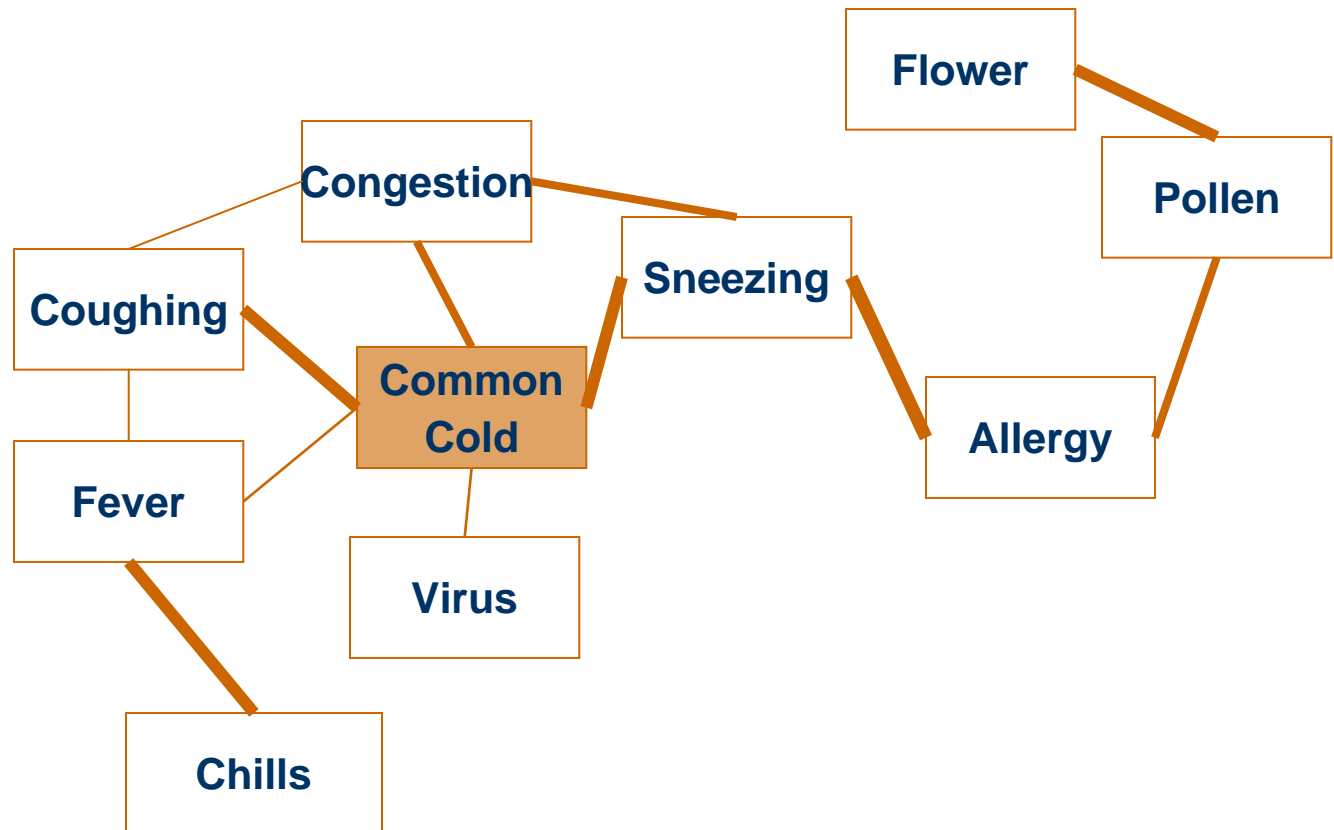


# “Concept” Search: Analyzing Word Relationships



[http://www.theseconconference.org/content/miscFiles/Best\\_Practices\\_Retrieval\\_Methods\\_revised\\_cover\\_and\\_preface.pdf](http://www.theseconconference.org/content/miscFiles/Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf)

# “Concept” Search: Network Effects

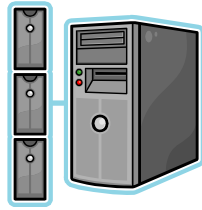


[http://www.theseconconference.org/content/miscFiles/Best\\_Practices\\_Retrieval\\_Methods\\_revised\\_cover\\_and\\_preface.pdf](http://www.theseconconference.org/content/miscFiles/Best_Practices_Retrieval_Methods_revised_cover_and_preface.pdf)

## Human Review (Computer Assisted)



Selection of keywords



Application of keywords



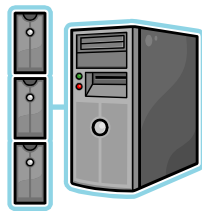
Human review of computer results (use coding manuals)



## Human Review (Computer Assisted)



Selection of  
keywords



Application of  
keywords



Human review of computer results (use coding manuals)

Substantial human  
involvement

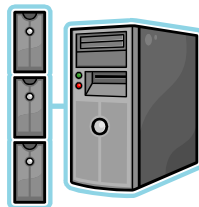


Lawyers  
Subject matter experts  
Linguists  
Statisticians

Iterative document  
sampling, analysis, and  
search refinement



Consistent replication  
of human judgments



## Computer Automation (Human Judgment)

- ✓ Today's search and review challenges in e-discovery
- ✓ Overview of search and review methods and technologies
- **Evolving perspectives from the bench**
  - Overview of benchmarking initiatives (TREC)
  - Q&A

- ***United States v. O’Keefe***  
537 F. Supp. 2d 14 (D.D.C. 2008)
- ***Equity Analytics v. Lundin***  
248 F.R.D. 331 (D.D.C. 2008)
- ***Victor Stanley, Inc., v. Creative Pipe, Inc.***  
250 F.R.D. 251 (D. Md. 2008)
- ***Mancia v. Mayflower Textile Services Co.***  
2008 U.S. Dist. LEXIS 83740 (D. Md. Oct. 15, 2008)

“Whether search terms or ‘keywords’ will yield the information sought is a complicated question involving the interplay, at least, of the **sciences of computer technology, statistics and linguistics** ... Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread.”

– U.S. Magistrate Judge John Facciola  
*United States v. O’Keefe*

(Emphasis added)

“[W]hile it is universally acknowledged that keyword searches are useful tools for search and retrieval of ESI, all keyword searches are not created equal; and there is a growing body of literature that highlights the risks associated with conducting an unreliable or inadequate keyword search or relying on such searches for privilege review.”

– U.S. Magistrate Judge Paul W. Grimm  
*Victor Stanley, Inc., v. Creative Pipe, Inc.*

# Guidance from Recent Court Opinions: TREC

“[The Text REtrieval Conference (TREC) LegalTrack] project can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search ...”

– U.S. Magistrate Judge Paul W. Grimm  
*Victor Stanley, Inc., v. Creative Pipe, Inc.*

# Guidance from Recent Court Opinions: The Sedona Conference

“...compliance with the Sedona Conference Best Practices for use of search and information retrieval will go a long way towards convincing the court that the method chosen was reasonable and reliable ...”

– U.S. Magistrate Judge Paul W. Grimm  
*Victor Stanley, Inc., v. Creative Pipe, Inc.*

- ✓ Today's search and review challenges in e-discovery
- ✓ Overview of search and review methods and technologies
- ✓ Evolving perspectives from the bench
- Overview of benchmarking initiatives (TREC)
- Q&A

- Sponsored by National Institute of Standards and Technology (NIST)
- Designed to evaluate effectiveness of search technologies in a real world legal context
- Open to all scientific research teams, e-discovery providers, law firms, etc.
- 9 hypothetical complaints and 80+ “requests to produce” or “topics” drafted by Sedona Conference members
- “Boolean negotiations” conducted as a baseline for search efforts
- Documents drawn from a publicly available 7 million-document tobacco litigation Master Settlement Agreement database

[http://www.thesedonaconference.org/content/miscFiles/TREC\\_OPEN\\_Letter.pdf](http://www.thesedonaconference.org/content/miscFiles/TREC_OPEN_Letter.pdf)

# TREC Legal Track Process

1. Complaints, including requests to produce (“topics”), are released to public
2. Participants clarify topic definitions with senior attorney TREC “Topic Authorities”
3. Participants employ respective methodologies to retrieve relevant documents
4. Participants deliver results to TREC Legal Track
5. Document reviewers review samples of results
6. Topic Authorities make final adjudication of disputed sample assessments
7. Results made public

<http://trec-legal.umiacs.umd.edu/2008InteractiveGuidelines.pdf>

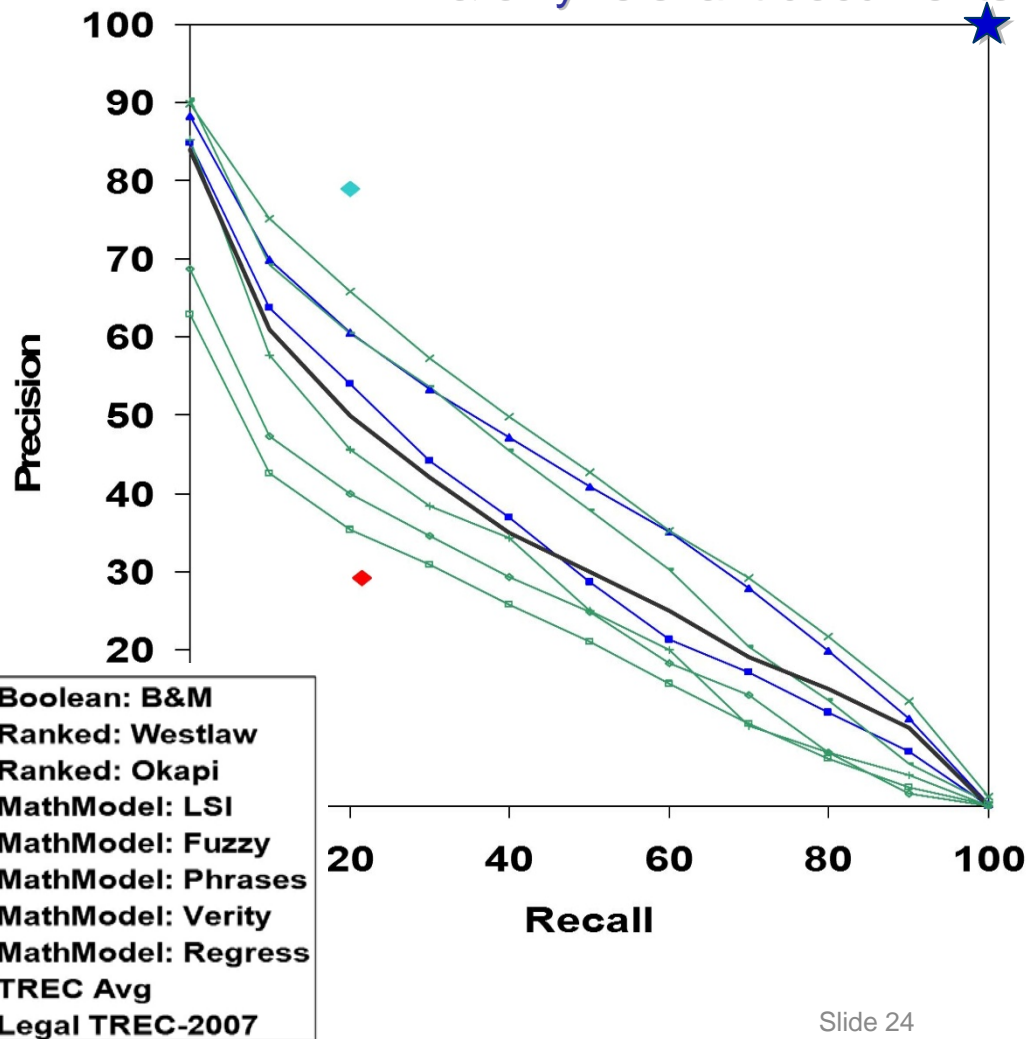
# The Sedona Conference Practice Points (Paraphrased)

1. In some cases, automated search methods may be reasonable, valuable, and necessary
2. Importance of well-thought-out process with substantial human input on front end
3. Choice of method tailored to specific legal context of case
4. Due diligence in selection of method
5. No guarantee that all responsive documents will be found, due to human language complexities; importance of measurement
6. Good faith effort to collaborate on use of search and retrieval methods
7. Expect to be able to explain choice of methodology
8. Be alert to new and evolving search and information retrieval methods

*The Sedona Conference Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery, 8 Sedona Conf. J. 189 (2007)*

# TREC Benchmark Measurements (circa 2007)

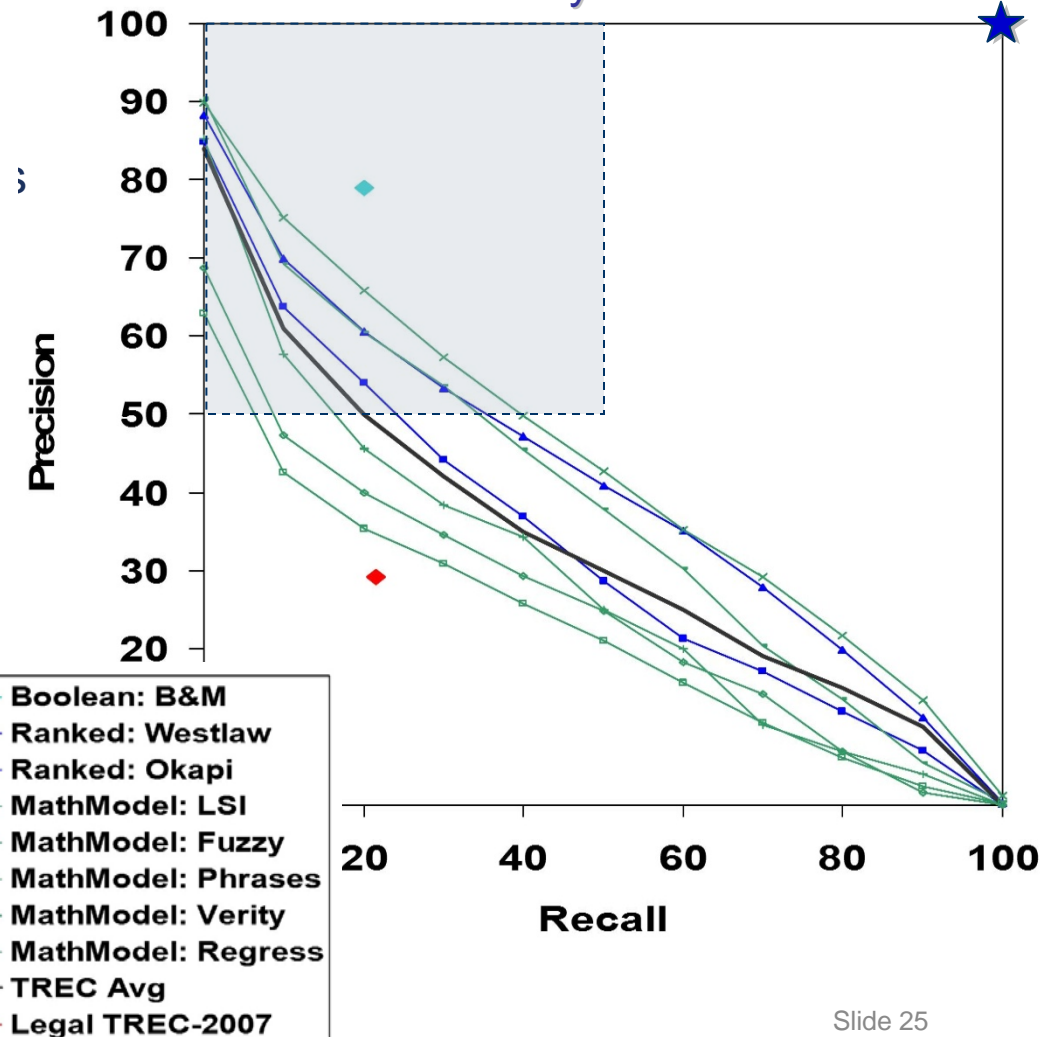
Find every relevant document  
& only relevant documents



## Computer Search:

- When over  $\frac{1}{2}$  of docs returned are relevant, less than  $\frac{1}{2}$  of all relevant docs are retrieved

Find every relevant document  
& only relevant documents

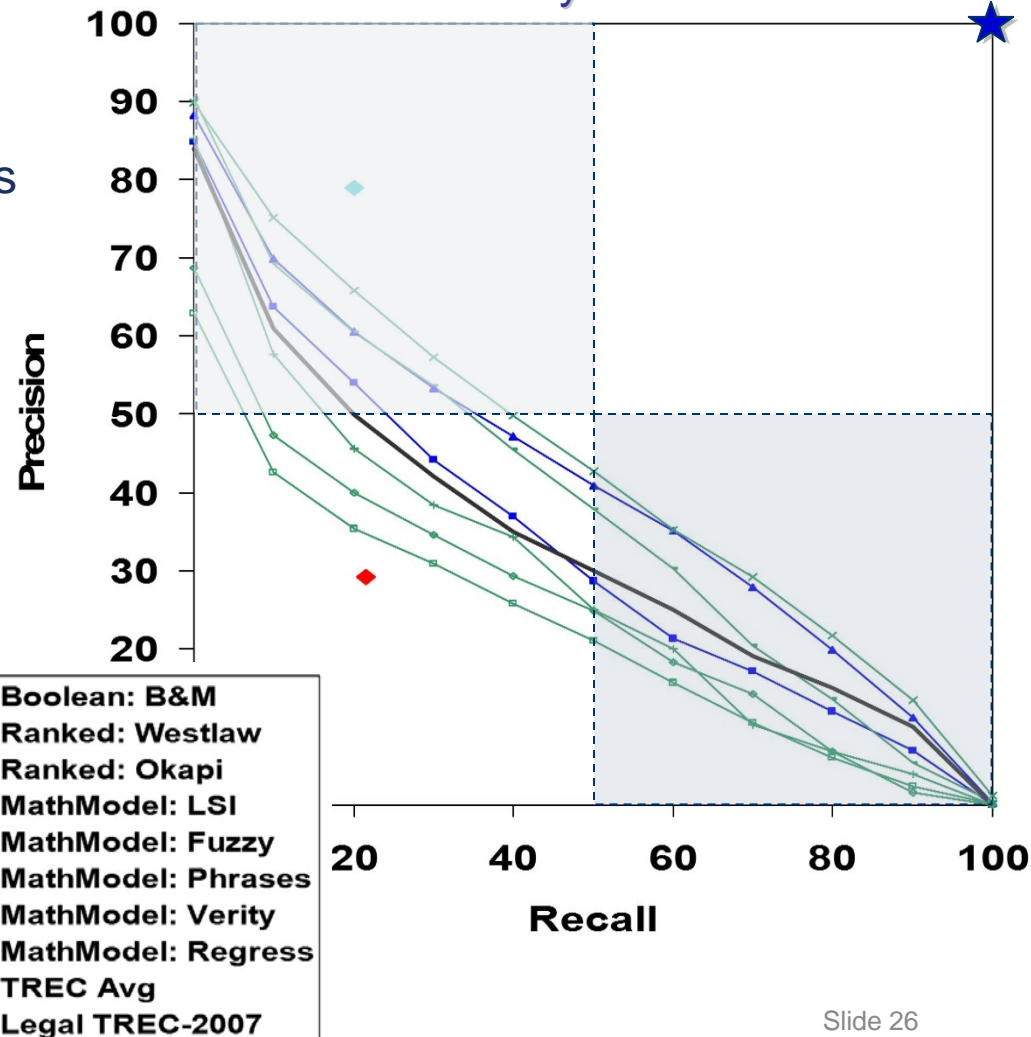


# TREC Benchmark Measurements (circa 2007)

## Computer Search:

- When over  $\frac{1}{2}$  of docs returned are relevant, less than  $\frac{1}{2}$  of all relevant docs are retrieved
- When over  $\frac{1}{2}$  of all relevant docs are retrieved, over  $\frac{1}{2}$  of docs returned are not relevant

Find every relevant document  
& only relevant documents



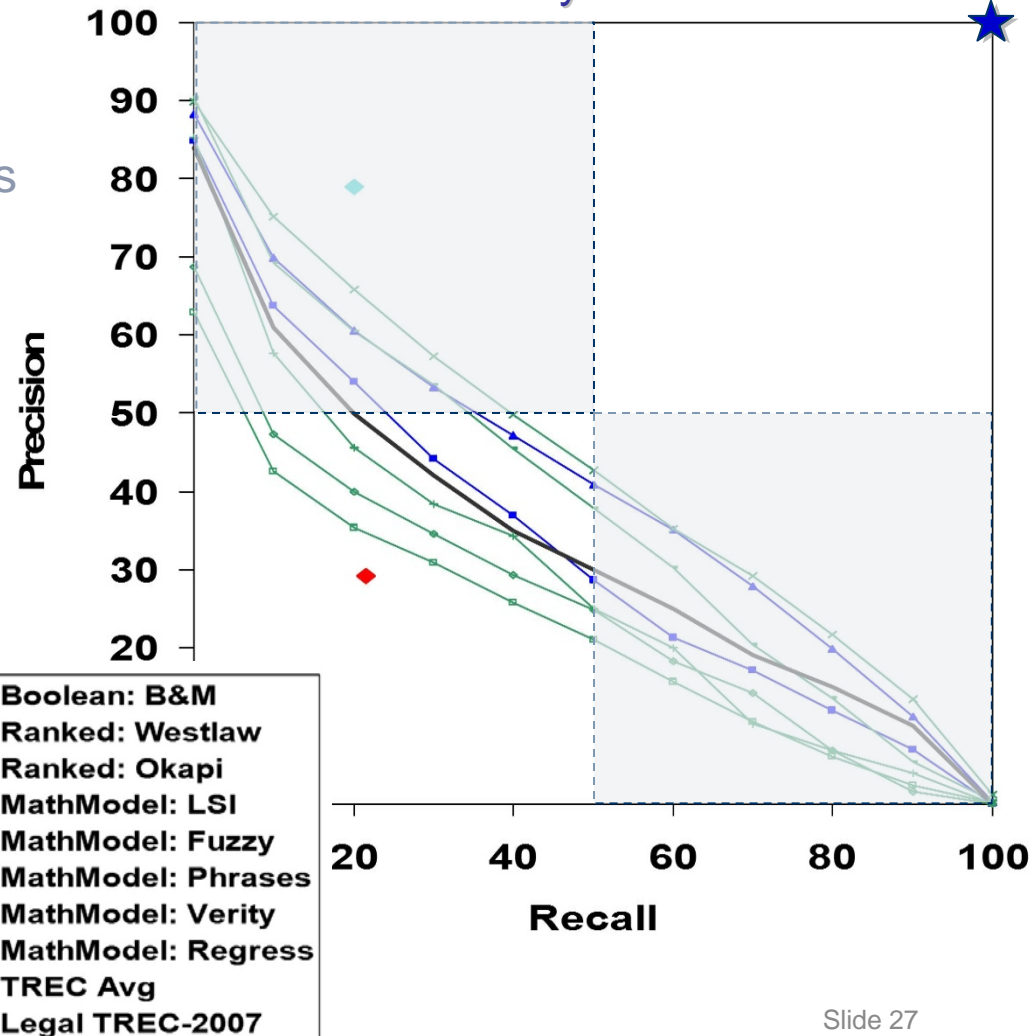
# TREC Benchmark Measurements (circa 2007)

## Computer Search:

- When over  $\frac{1}{2}$  of docs returned are relevant, less than  $\frac{1}{2}$  of all relevant docs are retrieved
- When over  $\frac{1}{2}$  of all relevant docs are retrieved, over  $\frac{1}{2}$  of docs returned are not relevant

Human Review less consistent than Automated Search  
(computer-assisted or not)

Find every relevant document  
& only relevant documents



# TREC Benchmark Measurements (circa 2007)

## Computer Search:

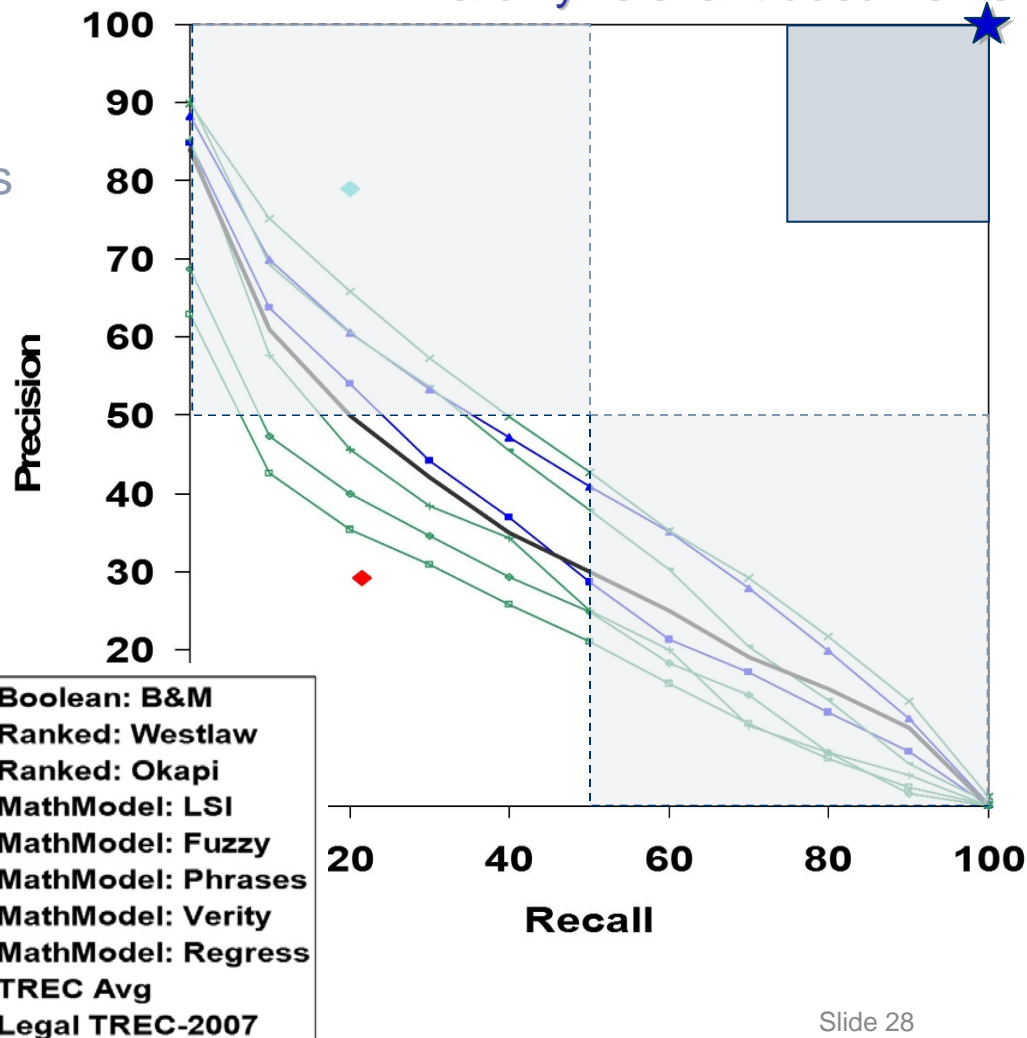
- When over  $\frac{1}{2}$  of docs returned are relevant, less than  $\frac{1}{2}$  of all relevant docs are retrieved
- When over  $\frac{1}{2}$  of all relevant docs are retrieved, over  $\frac{1}{2}$  of docs returned are not relevant

Human Review less consistent than Automated Search  
(computer-assisted or not)

## Human-assisted Computer Automation

- At least  $\frac{3}{4}$  of docs returned are relevant and at least  $\frac{3}{4}$  of all relevant docs are retrieved
- Cost is lower than computer-assisted human review

Find every relevant document  
& only relevant documents



- ✓ Today's search and review challenges in e-discovery
- ✓ Overview of search and review methods and technologies
- ✓ Evolving perspectives from the bench
- ✓ Overview of benchmarking initiatives (TREC)
- Questions?

- E-mail [info@h5.com](mailto:info@h5.com)
- Visit [www.h5.com](http://www.h5.com)

# For More Information on TREC Legal Track

- Download the TREC/Sedona Conference open letter:  
[www.thesedonaconference.org/content/miscFiles/TREC\\_OPEN\\_Letter.pdf](http://www.thesedonaconference.org/content/miscFiles/TREC_OPEN_Letter.pdf)
- Visit <http://trec-legal.umiacs.umd.edu/>
- Contact:
  - Jason R. Baron
  - Director of Litigation
  - Office of the General Counsel
  - National Archives and Records Administration
  - 8601 Adelphi Road, Suite 3110
  - College Park, MD 20740
  - tel. 301-837-1499 (direct)
  - 301-837-1750 (main)
  - 301-837-0293 (fax)
  - email: [jason.baron@nara.gov](mailto:jason.baron@nara.gov)