FUNDAMENTALS OF E-DISCOVERY

- for -

The Federal Practice Committee of the United States District Court for the District of Nebraska

* * *

January 7, 2010
University of Nebraska College of Law, Lincoln NE
Creighton University School of Law, Omaha NE

Maura R. Grossman
Wachtell, Lipton, Rosen & Katz

Today's Agenda

- **What is "ESI" and What Makes It So Different From Paper?**
- **A Brief Overview of Recent Case Law and Developments in E-Discovery:**
  - **Why Do You Need an E-Discovery Plan?**
    - The Challenge of E-Discovery: A "Perfect Storm;"
    - The Federal Rules of Civil Procedure.
  - **What are the Consequences of Getting E-Discovery Wrong?**
    - Lessons on Sanctions from 2009 and 2010.
  - **In 2011, What Will Courts Expect Regarding:**
    - Preservation and Collection;
    - Cooperation, Transparency, and the Meet-and-Confer Process;
    - Requests for Production, and Responses and Objections;
    - Reasonableness and Proportionality;
    - Form of Production.
- **Improving on Search and Review Using Sampling and Advanced Search Technology.**

What is "ESI"?

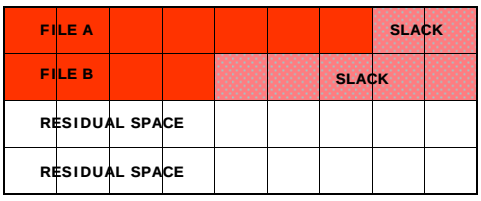
- "ESI" = **Electronically Stored Information**.
- Fed. R. Civ. P. 34(a)(1)(A): "writings, drawings, graphs, charts, photographs, sound recordings, images, and other **data or data compilations – stored in any medium from which information can be obtained** either directly or, if necessary, after translation by the responding party into a reasonably usable form[.]"
- Advisory Committee on the 2006 Amendments to the Federal Rules: "Rule 34 applies to **information that is fixed in a tangible form and** to information that is **stored in a medium from which it can be retrieved** and examined."

How ESI Differs from Paper

- ESI is easily and cheaply stored.
- ESI is capable of taking many forms and resides in multiple locations.
- ESI is voluminous and widely distributed.
- ESI is persistent yet fragile.
  - *** Delete ≠ Delete ***

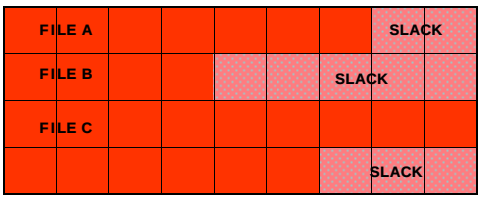
How Deletion and Overwriting Work

- Along with "active" files, a disk (or hard drive) contains "inactive" space, including both slack and residual space.



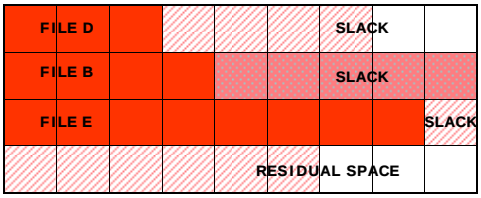
How Deletion and Overwriting Work (Continued)

- Let's assume that a new file (File C) is added to the "inactive" (residual) space on a disk or hard drive.



How Deletion and Overwriting Work (Continued)

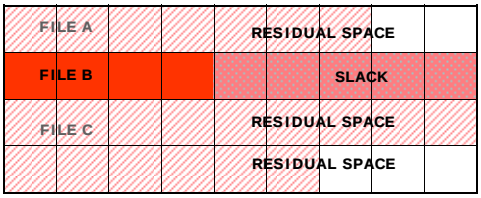
- If Files A and C are "deleted" and Files D and E are added, Files D and E are written over the spaces that Files A and C previously occupied, but they may not completely overwrite them. If Files D and E are smaller than Files A and C, portions of Files A and C will still remain in the slack space.



7

How Deletion and Overwriting Work (Continued)

- Therefore, "deleting" Files A and C does not completely remove them from the disk or hard drive; it simply allows those spaces to be overwritten by new files if and when that occurs.



8

How ESI Differs from Paper

- ESI is easily and cheaply stored.
- ESI is capable of taking many forms and resides in multiple locations.
- ESI is voluminous and widely distributed.
- ESI is persistent yet fragile.
- ESI is often created and maintained in complex, dynamic systems.
- ESI contains non-apparent information called "metadata."

9

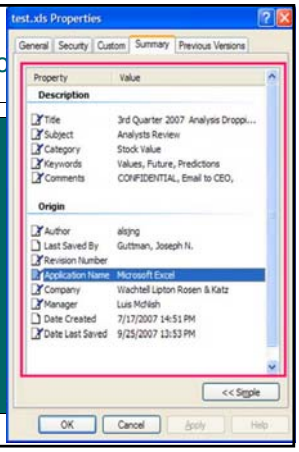
Types o

System Metadata

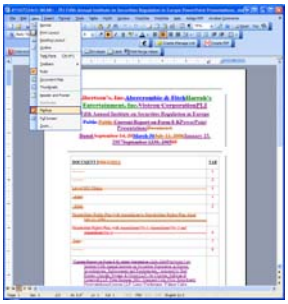
- **Stored separately from the underlying ESI.**
- **Assists the computer or network in storage and retrieval.**

Examples

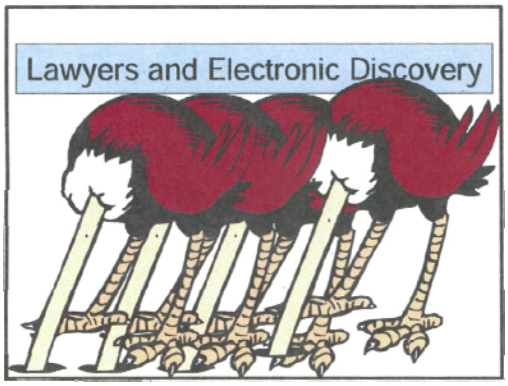
- **File name**
- **File size**
- **File location**
- **Author**
- **Date created**
- **Date last modified**



An Example of Metadata Gone Awry



11



12

Why Do You Need an E-Discovery Plan?

The challenge of e-discovery: a "perfect storm"

- Exploding data volumes;
- Multiplication of data sources and locations;
- New issues involving possession, custody and control;
- Challenges posed by cross-border e-discovery;
- Shorter deadlines and growing budgetary constraints;
- Litigators' tendencies to be reactive in nature;
- Lawyers' lack of knowledge or competence with respect to e-discovery;
- A "Wild West" atmosphere in the vendor community; and
- Higher standards and decreased patience on the part of the courts (therefore, leading to more sanctions).

13

Planning and Quality Requirements in the Federal Rules

- Rule 26(f)(3)(c) – Conference of the Parties; **Planning for Discovery**
  - "A **discovery plan** must state the parties' views and proposals on . . . [a]ny issues about disclosure or discovery of electronically stored information."
- Rule 26(g)(1)(B) – **Signing** Disclosures and **Discovery Requests, Responses, and Objections**
  - "By signing, **an attorney or party certifies** that to the best of the person's knowledge, information, and belief formed **after a reasonable inquiry**: [the discovery request, response or disclosure] is **not interposed for an improper purpose**, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation; and **neither unreasonable or unduly burdensome or expensive**, considering the needs of the case, prior discovery in the case, the amount in controversy, and the importance of the issues at stake in the action."
- Rule 37(a)(4) – **Evasive or Incomplete** Disclosure, Answer, or **Response**
  - "[A]n evasive or **incomplete** disclosure, answer, or **response** must be treated as a **failure to** disclose, answer, or **respond**."

14

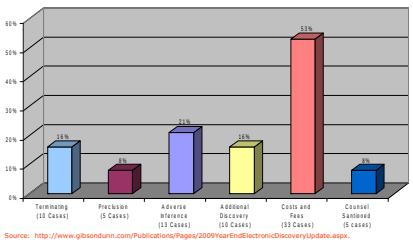
Consequences of Missteps in E-Discovery

- The number of **reported e-discovery opinions nearly doubled from 2008 to 2009**. That number **remained stable in 2010**.
- **In 2009, 42%** of the 208 e-discovery opinions **involved a motion for sanctions**. **In 2010, sanctions were sought** in slightly fewer cases (**38%** of the 209 opinions issued as of Dec. 1).
- **In 2009, sanctions** were **awarded in 70%** of those cases. **In 2010,** that number dropped somewhat to **62%**.
- Costs and fees were the most common sanction, but adverse inferences, preclusion of claims or defenses, and terminating sanctions (dismissal or default judgment) were all been awarded.
- **Counsel** were also **sanctioned individually**.
- Sanctioned conduct included **failure to preserve, search for, locate and/or produce** relevant, non-privileged evidence.
- The sanctioned conduct ranged from **negligent to bad faith**.

15

2009 E-Discovery Sanctions Cases

- There were 208 published e-discovery decisions in 2009 (and 209 as of Dec. 1, 2010).
- Sanctions were awarded in 62 of the 2009 decisions (and 49 of the 2010 decisions).
- More e-discovery sanctions were awarded in 2009 than in all prior years combined.



Source: http://www.gibsondunn.com/Publications/Pages/2009YearEndElectronicDiscoveryUpdate.aspx.

16

The Specter of Negligent Spoliation: *Residential Funding Corp.* v. *DeGeorge Fin. Corp.*

306 F.3d 99, 107, 108 (2d Cir. 2002).

Hon. José A. Cabranes

"[A] party seeking an adverse inference instruction based on the destruction of evidence must establish

(1) that the party having control over the evidence had an **obligation to preserve** it at the time it was destroyed;

(2) that the records were destroyed 'with a **culpable state of mind**'; and

(3) that the destroyed evidence was '**relevant**' to the party's claim or defense such that a reasonable trier of fact could find that it would support that claim or defense. . . ."

* * *

"[T]he '**culpable state of mind**' factor **is satisfied by a showing that the evidence was destroyed** 'knowingly, even without intent . . . or ***negligently***.'"

17

Selected Recent Opinions on Spoliation

- ***Victor Stanley, Inc.* v. *Creative Pipe, Inc.*, 250 F.R.D. 251 (D. Md. 2010) ("*Victor Stanley II*").**
- ***Rimkus Consulting Group* v. *Cammarata*, 688 F. Supp. 2d 598 (S.D. Tex. 2010).**
- ***Pension Comm. v. Banc of Am. Sec., L.L.C.*, 685 F. Supp. 2d 403 (S.D.N.Y. 2010).**
- *Triton Constr. Co.* v. *Eastern Shore Elec. Servs., Inc.*, 2009 WL 1387115 (Del. Ch. May 18, 2009), *aff'd* 988 A.2d 938 (Del. 2010).
- *Beard Research, Inc.* v. *Kates*, 981 A.2d 1175 (Del. Ch. 2009).
- *Green* v. *McClendon*, 262 F.R.D. 284 (S.D.N.Y. 2009).
- *Magaña* v. *Hyundai Motor Am.*, 167 Wash.2d 570 (2009).
- *Micron Tech, Inc.* v. *Rambus Inc.*, 255 F.R.D. 135 (D. Del. 2009).
- *Phillip A. Adams & Assocs., L.L.C.* v. *Dell, Inc.*, 621 F. Supp. 2d 1173 (D. Utah 2009).
- *Swofford* v. *Eslinger*, 671 F. Supp. 2d 1274 (M.D. Fla. 2009).
- *TR Investors, LLC* v. *Genger*, 2009 WL 4696062 (Del Ch. Dec. 9, 2009).
- *Doe* v. *Norwalk Cmty. College*, 248 F.R.D. 372 (D. Conn. 2007).
- *Phoenix Four, Inc.* v. *Strategic Res. Corp.*, 2006 WL 1409413 (S.D.N.Y. May 23, 2006).
- *United States* v. *Koch Indus., Inc.*, 197 F.R.D. 463 (N.D. Okla. 1998).

18

The Duty to Preserve

- The obligation typically arises when a party **knows or reasonably should know that evidence may be relevant to current or future litigation**, regardless of whether the party is the initiator or the subject of the litigation, or whether the context is a current or future regulatory investigation or proceeding.
- **Sources of the obligation to preserve:**
  - Common law duty to avoid spoliation;
  - Inherent power of the courts;
  - Court rules such as Fed. R. Civ. P. 37; and
  - Statutes or regulations such as 18 U.S.C. § 1519 (§ 802 of the Sarbanes-Oxley Act).
- The duty is triggered by a **reasonable anticipation of litigation**, based on a **credible threat**, which **may arise before a suit is filed** (especially for a plaintiff, who controls the timing of the litigation).
- A good rule of thumb is to consider whether you intend to claim **attorney work product** over documents prepared "in anticipation of litigation." (*See Siani* v. *SUNY at Farmingdale*, 2010 WL 3170664, at *5 (E.D.N.Y. Aug. 10, 2010).

19

The Duty to Preserve (Continued)

The assessment of **when** the duty to preserve is triggered is based on a **reasonable and good faith evaluation of the facts and circumstances known at the time**, including, but not limited to:

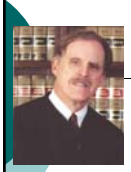
- **The level of knowledge** within the company **about the claim**, for example:
  - Who is making the claim or threat;
  - The nature and specificity of the claim or threat; and
  - Whether there have been similar claims against the company or others in the same industry.
- **The risk** to the company **of the claim**, for example:
  - The strength of the potential claim; and
  - How well known the claim is to those who might assert it, or to others.
- **The risk of losing information** if a litigation hold is not implemented, for example:
  - The likelihood that data could be lost or destroyed;
  - The significance of the data to the known or reasonably anticipated issues in dispute; and
  - The number and complexity of sources of information.

20

What is the Scope of the Duty to Preserve?

- Limited to **relevant information** under Fed. R. Civ. P. 26(b)(1).
- "Relevant" is broad, but not without limits.
- Scope may depend on, among other things:
  - The **issues raised** in the matter;
  - **Experience** in similar circumstances;
  - The **accessibility of the information**; and
  - The balance of **benefits and burdens** in preserving the information, including considerations of whether:
    - The information is **unique** or **critical to resolving the dispute**;
    - The data can be preserved at a **reasonable cost**;
    - Preservation would **disrupt routine activities or operations**; and
    - The requesting party would be **unfairly prejudiced** if the data is not preserved.
- There is no "one-size-fits-all" when it comes to preservation.

21



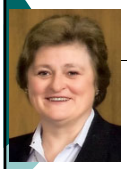
United States v. *Koch Industries, Inc.*

197 F.R.D. 463, 466, 481-82, 484 (N.D. Okla. 1998).

Hon. Sam A. Joyner

Lawsuit was filed in September 1991, but spoliation sanctions were imposed because defendant "**should have reasonably anticipated litigation**" as early as October 1986 – **nearly five years earlier** – because substantially similar allegations were made in litigation that was already pending. Sanctionable conduct included not only destruction of documents and ESI, but the failure to "develop[] or implement[] formal, company-wide information retention policies relating to the preservation of information for pending litigation or for other reasons."

22



Doe v. *Norwalk Community College*

248 F.R.D. 372, 377 (D. Conn. 2007).

Hon. Janet C. Hall

Defendant's **duty to preserve** arose:

- **Before** it received discovery requests;
- **Before** the lawsuit was filed;
- **"No later than"** when it received plaintiff's demand letter; and
- Probably **seven months earlier**, when defendant held an **internal meeting indicating** its **awareness of plaintiff's allegations**.

23

Phillip M. Adams & Assocs., L.L.C. v. *Dell, Inc.*

621 F. Supp. 2d 1173, 1191 (D. Utah 2009).

Hon. David Nuffer

Rejecting defendant's contention that the duty to preserve arose upon receipt of a demand letter in 2005, the court held that **"counsel's letter is not the inviolable benchmark,"** and that, because of earlier patent infringement litigation involving other parties, **defendant should have been sensitive to what was happening in the industry**. "Throughout this entire time, computer and component manufacturers were sensitized to the issue. . . . In the 1999-2000 environment, [defendant] should have been preserving evidence."

24

Pension Comm. v. *Banc of Am. Sec., L.L.C.*

685 F. Supp. 2d 456, 465 (S.D.N.Y. 2010).

Hon. Shira A. Scheindlin

"[T]he following failures support a finding of ***gross negligence***, when the duty to preserve has attached:

- to **issue a *written* litigation hold;**
- to **identify all of the key players** and to ensure that their electronic and paper records are preserved;
- to **cease the deletion of email** or to **preserve the records of former employees** that are in a party's possession, custody, or control; and
- to **preserve backup tapes when they are the sole source of relevant information or when they relate to key players**, if the relevant information maintained by those players is not obtainable from readily accessible sources."

25

What Must You Do When the Duty to Preserve is Triggered?

Pension Comm.* v. *Banc of Am. Sec., L.L.C., 685 F. Supp. 2d 456, 466, 473 (S.D.N.Y. 2010).

Hon. Shira A. Scheindlin

"[O]nce a party reasonably anticipates litigation, it must suspend its routine document retention/destruction policy and put in place a 'litigation hold' to ensure the preservation of relevant documents."

The litigation hold should "direct employees to *preserve* all relevant records – both paper and electronic . . . [and] create a mechanism for *collecting* the preserved records so that they can be **searched by someone other than the employee**."

26

The Risks of "Self Collection"

Roffe* v. *Eagle Rock Energy GP, L.P. and Eagle Rock Energy Partners, L.P., C.A. No. 5258-VCL (Del. Ch. 2010) (April 4, 2010 Transcript of Telephone Conference on Discovery Dispute, at pp. 9-10).

Hon. J. Travis Laster

The Court: "Am I correct that you have been relying on, for the other two committee members, what they **self-selected** to put in their transaction files, in terms of what you obtained and produced?"

Defense Counsel: "That's correct, Your Honor. . . ."

The Court: "Then here is my ruling. **This is not satisfactory**. . . . First of all, **you do not rely on a defendant to search their own e-mail system**. Okay? **There needs to be a lawyer who goes and makes sure the collection is done properly** . . . **we don't rely on people who are defendants to decide what documents are responsive**, at least not in this Court. And you certainly need to put someone on a plane to go out and see. . . ."

27

Some "Practice Points" on the Duty to Preserve

A Few Things to Consider:

- Information within the organization's "possession, custody, or **control**."
- Information held by **outsourced service providers**, storage facility operators and application service providers (*e.g.*, in the "cloud").
- Information created or maintained by **departing employees**.
- Information **located outside the U.S.**, especially in jurisdictions with privacy or data protection laws.
- **Monitoring** and **updating** the litigation hold.
- **Metadata** and **ephemeral data**.

28

Pension Comm. v. *Banc of Am. Sec., L.L.C.*

685 F. Supp. 2d 456, 461 (S.D.N.Y. 2010).

Hon. Shira A. Scheindlin

"In an era where vast amounts of electronic information is available for review, discovery in certain cases has become increasingly complex and expensive. **Courts cannot and do not expect that any party can meet a standard of perfection.** Nonetheless, the courts have a right to expect that litigants and counsel will take the necessary steps to ensure that relevant records are preserved when litigation is reasonably anticipated, and that such records are collected, reviewed, and produced to the opposing party."

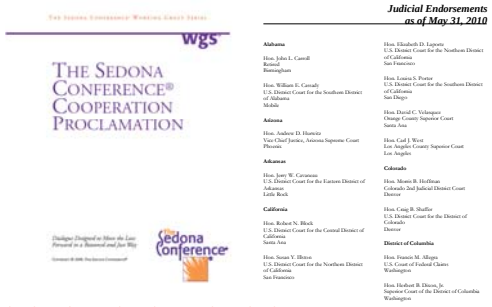
29

The Meet-and-Confer Process

- 2006 Rules amendments and subsequent case law urge – if not require – **greater cooperation and transparency**.
- Among the topics to be discussed at the meet-and-confer are:
  - Identification, preservation, collection, search and review of ESI (including custodians and data sources);
  - Confidentiality, privilege and other agreements (including FRE 502(e) agreements and **502(d) orders**, and **privilege logs**);
  - Technological challenges (including ESI that is "not reasonably accessible due to undue burden or cost"); and
  - Form of production.
- **Where appropriate, involve an e-discovery expert early** to help you develop a sound, defensible strategy.
- **Consider bringing your e-discovery expert to the meet-and-confer** session(s).

30

The Sedona Conference® Cooperation Proclamation



See http://www.thesedonaconference.org/content/tsc_cooperation_proclamation

Is Cooperation in Discovery an Oxymoron?

In re Seroquel Prods. Liab. Litig.

244 F.R.D. 650, 655, 660 (M.D. Fla. 2007).

Hon. David A. Baker

"Particularly in complex litigation, there is a heightened need for the parties to confer. . . .

"[T]he posturing and petulance displayed by both sides on this issue shows a disturbing departure from the expected professionalism necessary to get this case ready for appropriate disposition. **Identifying relevant records and working out technical methods for their production is a cooperative undertaking, not part of the adversarial give and take.**"

The Most Overlooked Rule: Fed. R. Civ. P. 26(g)

- 26(g)(1): A lawyer (or an unrepresented party) must sign every disclosure, discovery request, response or objection.
- 26(g)(1): The lawyer's signature certifies that "to the best of the person's knowledge, information, and belief formed after a reasonable inquiry;"
  - A disclosure is complete and correct; or
  - A discovery request, response or objection is
    - "**not interposed for any improper purpose, such as to** harass, cause unnecessary delay, or **needlessly increase the cost of litigation**; and
    - "**neither unreasonable nor unduly burdensome or expensive**. . . ."
- 26(g)(2): Unsigned disclosures, requests, responses and objections may be stricken.
- 26(g)(3): "If a certification violates this rule without substantial justification, the court . . . **must** impose an appropriate sanction."
  - Sanctions may include expenses and attorney's fees.

Mancia v. *Mayflower Textile Servs. Co.*

253 F.R.D. 354, 357-58 (D. Md. 2008).

Hon. Paul W. Grimm

Rule 26(g) "is intended to impose an 'affirmative duty' on counsel to behave responsibly during discovery, and to ensure that it is conducted in a way that is consistent 'with the spirit and purposes' of the discovery rules . . . which **requires cooperation rather than contrariety, communication rather than confrontation.**"

Rule 26(g) "aspires to eliminate . . . **kneejerk discovery requests served without consideration of cost or burden to the responding party** . . . [and] to bring an end to the equally abusive practice of **objecting to discovery requests reflexively – but not reflectively – and without a factual basis.**"

Mancia v. *Mayflower Textile Servs. Co.*

253 F.R.D. 354, 359 (D. Md. 2008).

Hon. Paul W. Grimm

"It would be difficult to dispute the notion that **the very act of making** such **boilerplate objections is *prima facie* evidence of a Rule 26(g) violation**, because if the lawyer had paused, made a reasonable inquiry, and discovered facts that demonstrated the burdensomeness or excessive cost of the discovery request, he or she should have disclosed them in the objection, as **both Rule 33 and 34 responses must state objections with particularity**, on pain of waiver."

Beard Research, Inc. v. *Kates*

981 A.2d 1175, 1187 (Del. Ch. 2009).

Hon. Donald F. Parsons, Jr.

"These realities [of electronic discovery] counsel strongly in favor of **early and**, if necessary, **frequent communications among counsel for opposing litigants to determine how discovery of ESI will be handled**. To the extent counsel reach agreements recognizing and permitting routine destruction of certain types of files to continue during litigation, the Court has no reason to object. Conversely, **if the parties do not focus on the handling of e-discovery in the early stages of a case, the Court is not likely to be sympathetic**. . . ."

37

Cartel Asset Mgmt. v. *Owcen Fin. Corp.*

2010 WL 502721, at *13-14 (D. Colo. Feb. 8, 2010).

Hon. Craig B. Shaffer

"Rule 26(c)(1) requires the moving party to certify that they have in good faith conferred or attempted to confer with the opposing party in an effort to resolve the dispute without court action. The obligation to 'meet and confer' is no less important or mandatory in cases characterized by recurring or fractious discovery disputes. Similarly, **the 'meet and confer' requirement should not be overridden by counsels' decision to approach discovery as a war of attrition**. . . . **Civil litigation**, particularly with the advent of expansive e-discovery, **has simply become too expensive and too protracted to permit superficial compliance with the 'meet and confer' requirement** under Rules 26(c) and 37(a)(1) and (d)(1)(B). . . . This court has endorsed *The Sedona Conference Cooperation Proclamation*. . . . **Counsel are on notice that, henceforth, this court will expect them to confer in good faith and make reasonable efforts to work together consistent with well-established case law and the principles underlying *The Cooperation Proclamation*.**"

38

The Most Under-Utilized Rule: Fed. R. Civ. P. 26(b)(2)(C)

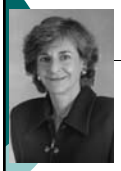
"On motion or on its own, the **court must limit** the frequency or extent of **discovery . . . if** it determines that:

(i) the discovery sought is unreasonably cumulative or duplicative, or can be obtained from some other source that is more convenient, less burdensome, or less expensive;

(ii) the party seeking discovery has had ample opportunity to obtain the information by discovery in the action; or

(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."

39



Rimkus Consulting Group, Inc. v. *Cammarata*

688 F. Supp. 2d 598, 613 (S.D. Tex. 2010).

Hon. Lee H. Rosenthal

"Whether preservation or discovery conduct is acceptable depends on what is ***reasonable***, and that in turn depends on whether what was done – or not done – was ***proportional*** to that case and consistent with clearly established applicable standards."

40

The Oft-Forgotten Rule: Fed. R. Civ. P. 34(b)

- 34(b)(1)(C) – The requesting party may specify the form of production of ESI.
- 34(b)(2)(D) – The responding party may object to the form specified; if so, it **must state the form it intends to use**.
  - If the requesting party disagrees, the parties must meet and confer.
  - If the parties cannot resolve the dispute, the requesting party may file a motion to compel.
- 34(b)(2)(D) – If the request does not specify a form of production, the responding party may choose, and **must state, the form it intends to use**.
- 34(b)(2)(E)(ii) – The responding party may select:
  - The form in which the material "is **ordinarily maintained**," or "a **reasonably usable form**."

41

Fed. R. Civ. P. 34(b)(2)(E)(i)

- Documents in a production may be produced either:
  - As they are kept in the **usual course of business**:
    - *See SEC* v. *Collins & Aikman Corp.*, 256 F.R.D. 403, 411-13 (S.D.N.Y. 2009) (Scheindlin, J.) (defining that as "a well-organized record-keeping system," and rejecting the SEC's argument that it could produce a "complete, unfiltered, and unorganized investigatory file");

    – or –

  - **Organized and labeled** to correspond to the categories in the request.

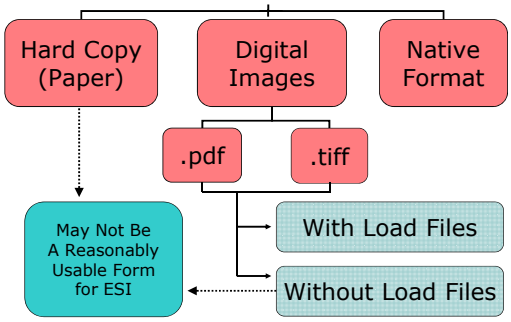
42

Form of Production: Technical Considerations

- What are the available forms of production?
  - What are image files (.tiff and .pdf)?
  - What are load files?
  - What is native form?
- What are the pros and cons of each form?
- What about form of production for structured data (*e.g.*, databases), SharePoints, and data in "the Cloud"?

43

Available Forms of Production



44

Basic Metadata Fields

Bates_Begin	The Bates number of the first page of a document
Bates_End	The Bates number of the last page of a document
Attach_Begin	The Bates number of the first page of the first document in a document family (*e.g.*, an email with one or more attachments)
Attach_End	The Bates number of the last page of the last document in a document family (*e.g.*, the last page of the last attachment)
Sent_Date	For an email, the sent date of the message
Sent_Time	For email, the sent time of the message
To	The recipient(s) of an email message, in a semi-colon delimited multi-value list
From	The sender of an email message
CC	The copy(s) of an email message, in a semi-colon delimited multi-value list
BCC	The blind copy(s) of an email message, in a semi-colon delimited multi-value list

45

Basic Metadata Fields

Subject/Title	The subject of an email, or the filename of an attachment or stand-alone e-file
Text	The body of an email, attachment or stand-alone e-file, either from OCR or extracted from the electronic file
Custodian	The custodian(s) in whose file(s) the document was found, in a semi-colon delimited multi-value list
Native_File	The file path to the location of the native version of the file being produced, if produced natively

46

Considerations in Choosing or Challenging a Form of Production

- The time and cost of processing;
- The ease and cost of search, review and use;
- The need for metadata and the extent to which metadata will enhance the usability of the ESI,
  - *See, e.g., Aguilar* v. *Immigration and Customs Enforcement Div. of the U.S. Dept. of Homeland Sec.*, 255 F.R.D. 350 (S.D.N.Y. 2008) (Mass, J.);
- The ability to Bates-number pages, mark individual pages as "confidential" and redact;
- The ability to maintain integrity and prove authenticity; and
- Compliance with Rule 34 (for example, a party may not "downgrade" the ability to search ESI).

47

Take-Aways on Rule 26

- **Meet and Confer** with your adversary about ESI early, often and in good faith, concerning, among other things:
  - Scope of preservation and production;
  - Accessibility and cost;
  - Form of production, clawback agreements, **privilege logs**, etc.
- Present the Court with a **reasonable** e-discovery **plan**.
  - Where appropriate, consult with an e-discovery expert about the plan.
- **Memorialize agreements** between the parties and incorporate them into court orders, **especially FRE 502(d) orders**.
- **Avoid** overly broad and onerous discovery requests, and formulaic or overly combative responses and objections.
- **Object** to such behavior by your adversary by using Rules 26(b)(2)(C) and 26(g).

48

Why is Search Important?

Suppose you have a document collection consisting of a billion emails, 25% of which have attachments. Assume a review rate of approximately 50 documents per hour. At this rate, it would take 100 contract attorneys, working 10 hour days, 7 days a week, 52 weeks a year, how long to complete?

APPROXIMATELY 55 YEARS TO COMPLETE!

- At a blended rate of $100 per hour, this review would cost **$2 billion**.
- Even if the document set was **reduced by 99%** (to 10 million emails), it would take almost **29 weeks** to review and would cost approximately **$20 million**.

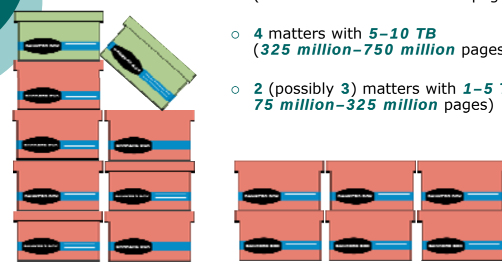
A. Eight months.

B. Nineteen months.

C. About four and a half years.

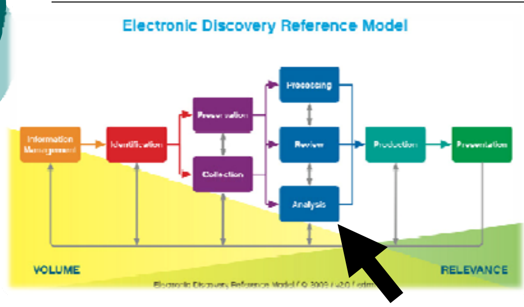
D. Over a decade.

Representative Matters from 2008 to 2010



- **2** matters with ***10–15 TB*** (***750 million–1.125 billion*** pages)
- **4** matters with ***5–10 TB*** (***325 million–750 million*** pages)
- **2** (possibly **3**) matters with ***1–5 TB*** ***75 million–325 million*** pages)

"The Half of Knowledge is Knowing Where to Find It." *– Samuel Johnson (1775)*



In re Seroquel Prods. Liab. Litig.

244 F.R.D. 650, 662 (M.D. Fla. 2007).

Hon. David A. Baker

"[W]hile key word searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process. . . . Common sense dictates that **sampling and other quality assurance techniques must be employed** to meet requirements of completeness."

United States v. *O'Keefe*

537 F. Supp.2d 14, 24 (D.D.C. 2008).

Hon. John M. Facciola

- "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**."
- "This topic is **clearly beyond the ken of a layman**. . . ."

Victor Stanley, Inc. v. *Creative Pipe, Inc.* ("*Victor Stanley I*")

250 F.R.D. 251 (D. Md. 2008).

Hon. Paul Grimm

- Creative Pipe's lawyers inadvertently produced 165 documents over which they later claimed privilege.
- Judge Grimm ruled that the **defendant had waived the privilege** because they:
  - **Did not** employ **search expertise**;
  - **Did not test the keyword list**;
  - **Did not QC the results** that were produced; and
  - **Did not** implement a **clawback agreement**.

Victor Stanley, Inc. v. *Creative Pipe, Inc. ("Victor Stanley I")*

250 F.R.D. 251, 256-57 (D. Md. 2008).

Hon. Paul Grimm

"[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 exclusively on such searches for privilege review."

55

William A. Gross Constr. Assocs. Inc. v. *American Mfrs. Mut. Ins. Co.*

256 F.R.D. 134, 134 (S.D.N.Y. 2009).

Hon. Andrew Peck

In a multi-million dollar dispute involving the production of non-party emails, where the parties could not agree on keyword search terms, the Court crafted a search methodology, stating that:

"This Opinion should serve as a **wake-up call to the Bar** . . . about the need for **careful thought, quality control, testing, and cooperation with opposing counsel in designing search terms** or 'keywords' to be used to produce emails or other electronically stored information ('ESI'). While this message has appeared in several cases from outside this Circuit, it appears that **the message has not reached many members of our Bar**."

56

William A. Gross Constr. Assocs. Inc. v. *American Mfrs. Mut. Ins. Co.*

256 F.R.D. 134, 135-36 (S.D.N.Y. 2009).

Hon. Andrew Peck

"This case is just the latest example of **lawyers designing keyword searches in the dark, by the seat of the pants**, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails."

"[W]here counsel are using keyword searches for retrieval of ESI, they **at a minimum must carefully craft the appropriate keywords, with input from the ESI's custodians** as to the words and abbreviations they use, and the proposed methodology **must be quality control tested** to assure accuracy in retrieval and elimination of 'false positives.' It is time that the Bar – even those lawyers who did not come of age in the computer era – understand this."

57

In re Fannie Mae Sec. Litig.

552 F.3d 814 (D.C. Cir. 2009).

Hon. David Tatel

Upholding contempt citation against government agency where it failed to meet extended, agreed-upon deadlines, after committing to produce all non-privileged documents responsive to **400 search terms**, which yielded a review set consisting of **660,000 documents** that needed to be, but could not be, reviewed in time, and cost **$6 million**, or **9% of the agency's annual budget** to review.

58

Smith v. *Life Investors Ins. Co. of Am.*

2009 WL 2045197 (W.D. Pa. July 9, 2009).

Hon. Terrence McVerry

After defendant selected and ran search terms without input from plaintiff, plaintiff moved to compel the list of terms used, and defendant claimed attorney work-product; court granted the motion, citing *Victor Stanley*, ruling that **defendant had a duty to demonstrate its search methodology was reasonable**, which could be done through identification of the **keywords used**, an explanation of the **qualifications of those selecting the keywords** and **proof of quality assurance testing**.

59

Pension Comm. v. *Banc of Am. Sec., L.L.C.*

685 F. Supp. 2d 456, 465 (S.D.N.Y. 2010).

Hon. Shira A. Scheindlin

Defining negligence, gross negligence and willfulness in the discovery context, Judge Scheindlin noted that "[r]ecent cases have also addressed . . . the failure to assess the accuracy and validity of selected search terms," and labeled that as "**negligence**."

60

Mt. Hawley Ins. Co. v. *Felman Prod., Inc.*

2010 WL 1990555, at *1, 12-13 (S.D. W. Va., May 18, 2010).

Hon. Mary E. Stanley

- Court held that privilege was waived for 377 inadvertently produced documents, in a massive "document dump," where 30% of the pages were irrelevant, and plaintiff failed to take sufficient precautions to prevent disclosure.
- The plaintiff "[p]roduced more than 346 gigabytes of data **without sampling for relevancy, over-inclusiveness or under-inclusiveness**," and "[m]arked all 346 gigabytes of data as 'CONFIDENTIAL.'"
- The plaintiff "**failed to perform critical quality control sampling** to determine whether their production was appropriate and neither over-inclusive nor under-inclusive. . . ."
- "[T]he **failure to test the reliability of keyword searches by appropriate sampling** is imprudent . . . [and] underscores the **lack of care taken in the review process**."

61

Challenges of Search and Retrieval

- Poor records management.
- Heterogeneous ESI presents an array of technical issues (*e.g.*, passwords, encryption and embedded objects).
- Non-English and non-textual forms of ESI (*e.g.*, audio, video, etc.).
- OCR issues.
- Inherent ambiguity of language:
  - Synonymy = variation in describing the same person or thing (*e.g.*, "diplomat," "consul," "official," "ambassador").
  - Polysemy = ambiguous terms (*e.g.*, "bush," "strike").

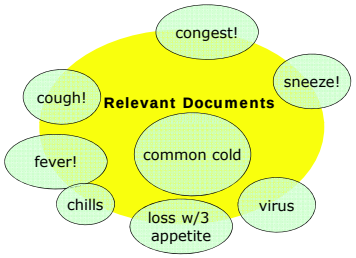
62

Challenges of Search and Retrieval (Continued)

- Ubiquity of human error (*e.g.*, misspellings, typos).
- Abbreviations, colloquialisms, slang, code words and new short forms used in text messaging and IM.
- Practical limits imposed by deadlines and resource constraints.
- Failure to understand and employ "best practices."
- Asymmetry in knowledge.

63

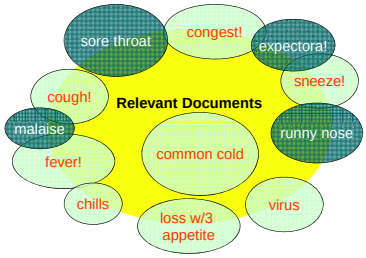
Keyword or "Boolean" Search



Source: http://www.thesedonaconference.org/content/miscFiles/Best_Practices_Retrieval_Methods___revised_cover_and_preface.pdf.

64

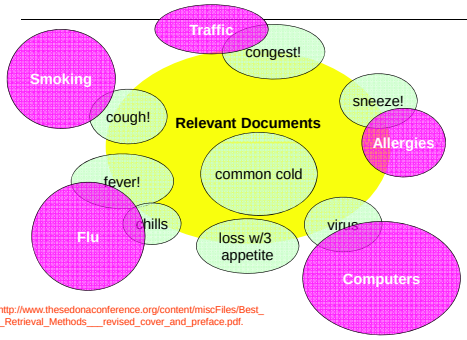
The Problem of Under-Inclusion



Source: http://www.thesedonaconference.org/content/miscFiles/Best_Practices_Retrieval_Methods___revised_cover_and_preface.pdf.

65

The Problem of Over-Inclusion



Source: http://www.thesedonaconference.org/content/miscFiles/Best_Practices_Retrieval_Methods___revised_cover_and_preface.pdf.

66

The Blair and Maron Study (1985)

- In a landmark 1985 study, Blair & Maron found a **disconnect between lawyers' perceptions** of their ability to find relevant documents **and their actual ability** to do so.
- In a case with 40,000 documents (350,000 pages), lawyers estimated that their manual search had identified **75%** of relevant documents, when in fact only about **20%** were found.

Source: David C. Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, Communications of the ACM, Vol. 28, Issue 3 (March 1985).

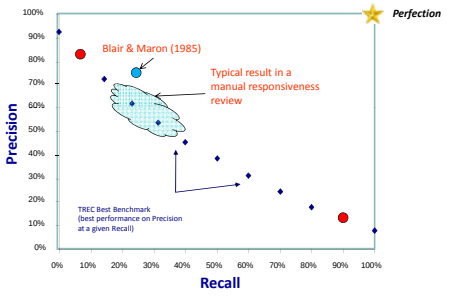
Measures of Information Retrieval

- **Recall** =
  # of responsive documents *retrieved*
  Total # of responsive documents *in the entire document collection*

  ("How many of the responsive documents did I find?")
- **Precision** =
  # of *responsive documents* retrieved
  Total # of *documents* retrieved

  ("How much of what I retrieved was junk?")

The Recall-Precision Trade-Off



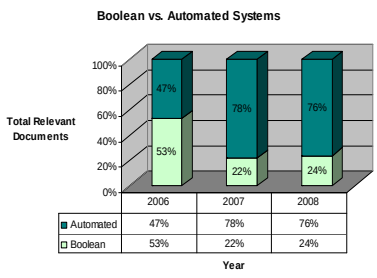
What Does Recent Research Tell Us: The Text REtrieval Conference (TREC)

- International, interdisciplinary research project sponsored by the National Institute of Standards and Technology (NIST), which is part of the U.S. Department of Commerce.
- Designed to promote research into the science of information retrieval.
- First TREC conference was held in 1992; the TREC Legal Track began in 2006.
- Designed to evaluate the effectiveness of search technologies in the context of e-discovery.
- Employs hypothetical complaints and requests for production drafted by members of The Sedona Conference®.
- Boolean negotiations were conducted as a baseline for search efforts.
- For the first three years (2006-2008), documents were from the publicly available 7 million document tobacco litigation Master Settlement Agreement database.
- In 2009, a publicly available Enron data set was used for the first time.
- Participating teams of information scientists from around the world and U.S. litigation support service providers have contributed computer runs attempting to identify responsive documents.

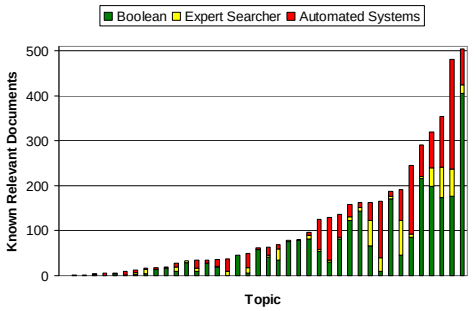
The TREC Interactive Task: Key Features

- Interactive Task was introduced in 2008.
- It models a document review for responsiveness.
- It begins with a mock complaint and associated requests for production ("topics").
- It has a single Topic Authority ("TA") for each topic.
- Teams may interact with the Topic Authority for up to 10 hours.
- Each team must submit a binary ("responsive"/"unresponsive") classification for each document in the collection for their assigned topic(s).
- It provides for a two-step assessment and adjudication process.

Boolean vs. Automated Systems: Results of the TREC Legal Track for Years 2006-2008



Nobody Finds Everything



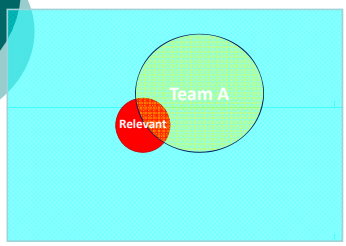
Source: TREC 2006 Legal Track

TREC Topic 204 (2009)

Document Request:

"All documents or communications that describe, discuss, refer to, report on, or relate to any intentions, plans, efforts, or activities involving the alteration, destruction, retention, lack of retention, deletion, or shredding of documents or other evidence, whether in hard-copy or electronic form."

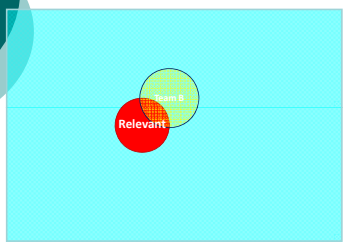
TREC Topic 204: Recall and Precision



Team A:
Search Term-Based Process

- Recall = 30.5%
- Precision = 7.7%
- F1 = 12.3%

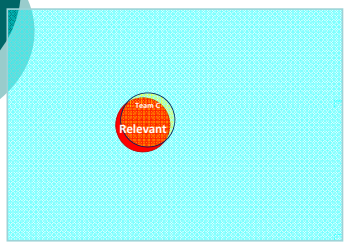
TREC Topic 204: Recall and Precision



Team B:
Independent Law Firm / Vendor Process

- Recall = 19.8%
- Precision = 16.9%
- F1 = 18.3%

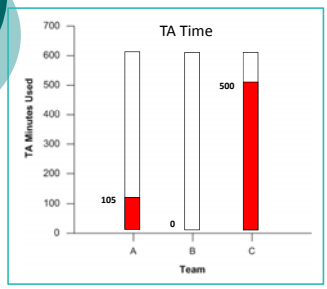
TREC Topic 204: Recall and Precision



Team C:
Technology-Assisted Collaborative and Iterative Process

- Recall = 76.2%
- Precision = 84.4%
- F1 = 80.1%

TREC Topic 204: TA Time and Precision



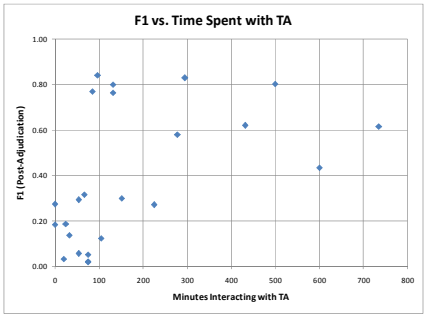
Submitted as Relevant
(out of 569,034 messages)

Team A:
12,748 Messages

Team B:
3,741 Messages

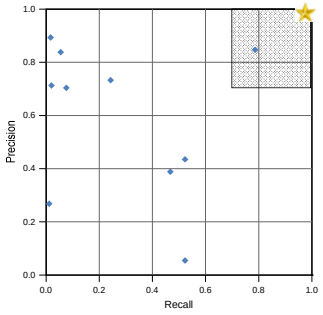
Team C:
2,919 Messages

Team-TA Interaction and Effectiveness



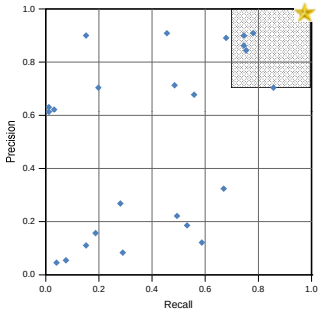
79

TREC Interactive Task – 2008 Results



80

TREC Interactive Task – 2009 Results



81

Suggestions for Improving Keyword Search

- Start with the document request:
  - Who are the key custodians?
  - What is the applicable time frame?
  - What terms-of-art are employed?
- Translate the request into plain English.
- Involve multiple people to get differing interpretations of the requests and potential keywords from different vantage points.
- Seek input from the people who actually created, sent or received the documents.
- Would a linguist or substantive expert be helpful?

82

Suggestions for Improving Keyword Search (Continued)

- Look at a sample of known responsive documents for unique words or phrases. In what context do those words or phrases appear?
- Consider obtaining an index from the vendor of all words and domain names in the document set and their frequencies from most to least frequent.
- Consider incorporating common misspellings, errors, variants and synonyms. (You can utilize tools on the web for this task.)
- Determine irrelevant file types and domain names that can be removed.
- Develop search strategies for handwritten and foreign-language documents, drawings, facsimiles, password-protected and/or encrypted files.

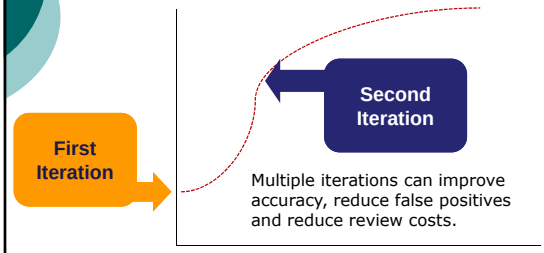
83

Suggestions for Improving Keyword Search (Continued)

- Know the capabilities and limitations of the available search tools and choose the most appropriate one for your particular search strategy.
  - When appropriate, consult with an e-discovery expert in the selection process.
- Take random samples of 400-700 documents and test your search terms.
  - Test from both the "hits" and "misses" piles.
  - Retest as necessary and appropriate under the circumstances.
- Keep track of and document what was done so you can later explain the rationale for the process or method you applied.
- Collaborate and communicate in good faith.

84

Improving Search Effectiveness Through Relevance Feedback



Multiple iterations can improve accuracy, reduce false positives and reduce review costs.

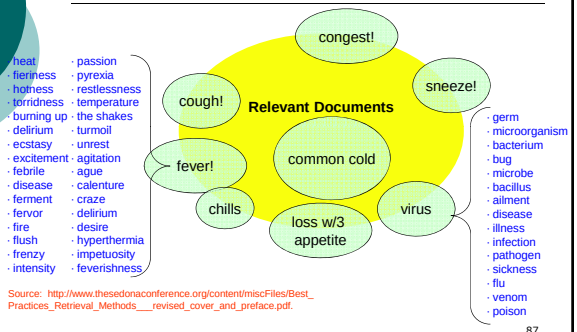
Source: F.C. Zhao, D.W. Oard, and J.R. Baron, *Improving Search Effectiveness in the Legal E-Discovery Process Using Relevance Feedback* (paper presented at the DESI III Global E-Discovery/E-Disclosure Workshop at the 12th International Conference on Artificial Intelligence and Law, 2009).

Beyond "Boolean": Alternative Search Methods

"Concept Search," Categorization and Ranked Retrieval Tools

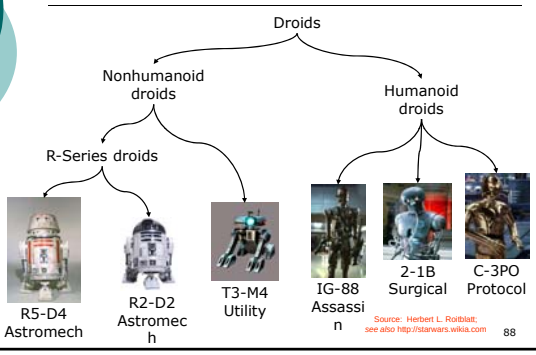
- Linguistic Methods:
  - Thesauri;
  - Taxonomies;
  - Ontologies.
- Statistical Methods:
  - Neural Networks;
  - Latent Semantic Indexing;
  - Machine Learning (a.k.a "predictive coding").

"Concept Search": Thesauri



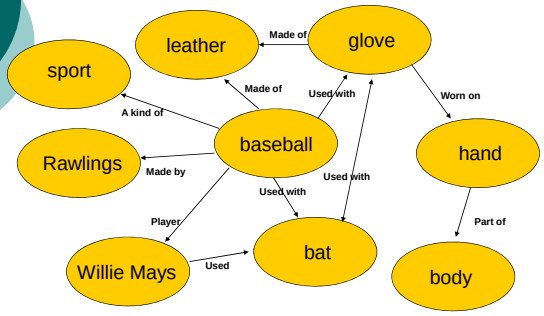
Source: http://www.thesedonaconference.org/content/miscFiles/Best_Practices_Retrieval_Methods___revised_cover_and_preface.pdf.

"Concept Search": Taxonomies

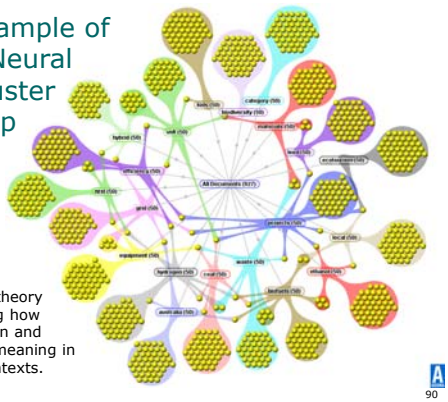


Source: Herbert L. Roitblatt; see also http://starwars.wikia.com

"Concept Search": Ontologies



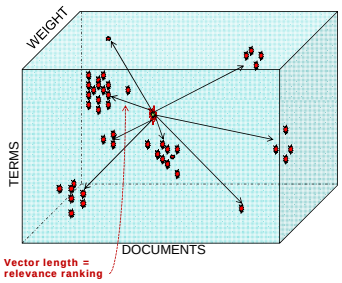
Example of a Neural Cluster Map



Based on a theory of replicating how humans learn and distinguish meaning in different contexts.

Latent Semantic Indexing

1. The system analyzes documents along certain dimensions and assigns each record a place by creating "clusters."
2. Query documents are analyzed and placed in the matrix.
3. "Hits" and rankings are determined by the distance from the clusters.



91

Defining Machine Learning (a.k.a. "Predictive Coding")

- The use of machine learning technologies to categorize an entire collection of documents as responsive or non-responsive, based on human review of only a subset of the document collection. These technologies typically rank the documents from most to least likely to be responsive to a specific information request. This ranking can then be used to "cut" or partition the documents into one or more categories, such as potentially responsive or not, in need of further review or not, etc.
  - Think of a spam filter that reviews and classifies email into "ham," "spam," and "questionable."

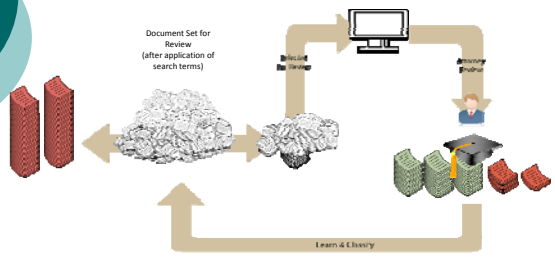
92

Types of Machine Learning

- **SUPERVISED LEARNING** = where a **human chooses the document exemplars** ("seed set") to feed to the system and requests that the system rank the remaining documents in the collection according to their similarity to, or difference from, the exemplars (*i.e.*, "find more like this").
- **ACTIVE LEARNING** = where the **system chooses the document exemplars** to feed to the human and requests that the human make responsiveness determinations from which the system then learns and applies that learning to the remaining documents in the collection.

93

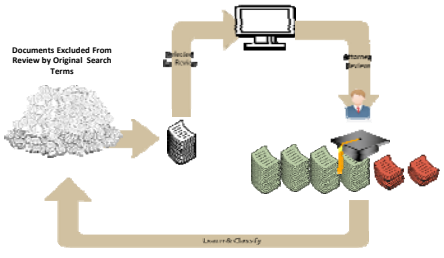
Machine Learning: Achieving High Precision



Source: Servient Inc. http://www.servient.com/

94

Machine Learning: Improving Recall



Source: Servient Inc. http://www.servient.com/

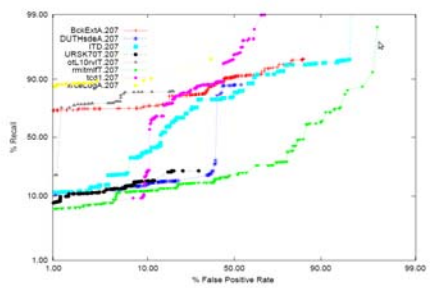
95

Results From the TREC 2010 Learning Task: Recall at 30% Retrieved

run	200	201	202	203	204	205	206	207	Avg.	Est.	Acc.
otL10rvlT	39.8	**85.5**	**100.2**	**100.5**	**85.2**	**84.6**	**98.9**	**86.6**	**85.1**	98.8	86.1
xrceLogA	**77.9**	**93.8**	**99.4**	**92.2**	73.8	71.8	**74.9**	**92.0**	**84.4**	88.8	95.0
DUTHsdtA	**90.6**	**85.0**	**97.8**	**103.4**	**98.0**	**82.2**	**88.0**	18.7	**82.9**	90.1	92.0
tcd1	67.2	61.6	**98.2**	**77.9**	**76.2**	57.5	**97.5**	**87.4**	**77.9**	55.8	71.6
rmitindA	72.9	**85.8**	**96.7**	**102.5**	**79.2**	**87.7**	**78.3**	19.8	**77.8**	53.3	68.5
BckExtA	**78.9**	**75.1**	**90.0**	38.6	67.5	72.7	**85.5**	**80.9**	73.6	49.7	67.5
ITD	N/A	45.6	67.5	20.7	41.6	35.2	29.7	74.7	44.9	61.8	72.6
URSK70T	51	17.7	18.6	23.8	62.2	22.4	**87.6**	22.6	38.2	9.01	42.0

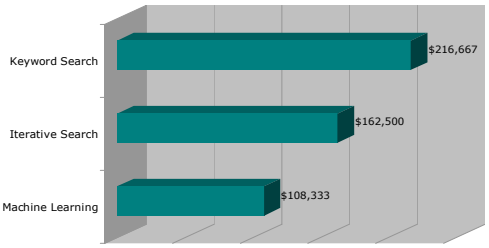
96

Results From the TREC 2010 Learning Task



97

Comparing the Cost-Savings Using Keyword Search, an Iterative Search Process, and Machine Learning



Better search also decreases the time needed for review

Source: Servient Inc. http://www.servient.com/

98

Disability Rights Council of Greater Wash. v. *Washington Metro. Transit Auth.*

242 F.R.D. 139, 148 (D.D.C. 2007).

Hon. John M. Facciola

"I bring to the parties' attention recent scholarship that argues that **concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results**. *See* George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?* 13 Rich. J.L. & Tech. 10 (2007)."

99

Take-Aways on Search and Retrieval

- Successful search requires *both* a sound process and the appropriate technology.
- There is no substitute for careful planning and informed legal judgment.
- No single search methodology has been shown to be significantly better in *all* circumstances.
- The use of multiple approaches in combination can improve the accuracy of results.
- An e-discovery expert can help to select the right tool and to develop a practical and defensible search protocol.
- It will be increasingly important to stay abreast of developments in search as the technology continues to evolve and improve.

100

Some Industry Resources to Watch

The Sedona Conference®
See www.thesedonaconference.org

- TREC – the **T**ext **RE**trieval **C**onference
  - 2010 Legal Track results are due out shortly (Feb. 2011).
  - 2011 Legal Track has been approved and design is underway.
    See http://trec-legal.umiacs.umd.edu/
- EDRM – Electronic Discovery Reference Model
  See www.edrm.net

101

Questions?

- Contact Information:
  - Maura R. Grossman
    Wachtell, Lipton, Rosen & Katz
    51 West 52nd Street, 31st Floor
    New York, New York 10019-6150
    (212) 403-1391
    MRGrossman@wlrk.com
- Thank you!

102