Inherent Problems in e-discovery and Solutions

Jianqing Wu, Ph.D.
Founder, A & U Data

A & U Data, v0.90, July 12, 2016

Network-Based Review Model

- The Network-based review model has been used for more than a decade. By using the review model, documents are collected and distributed to a group of reviewers for review.

- Several inherent problems directly affect review quality, protection of privileged documents, trade secrets, and other harmful information.

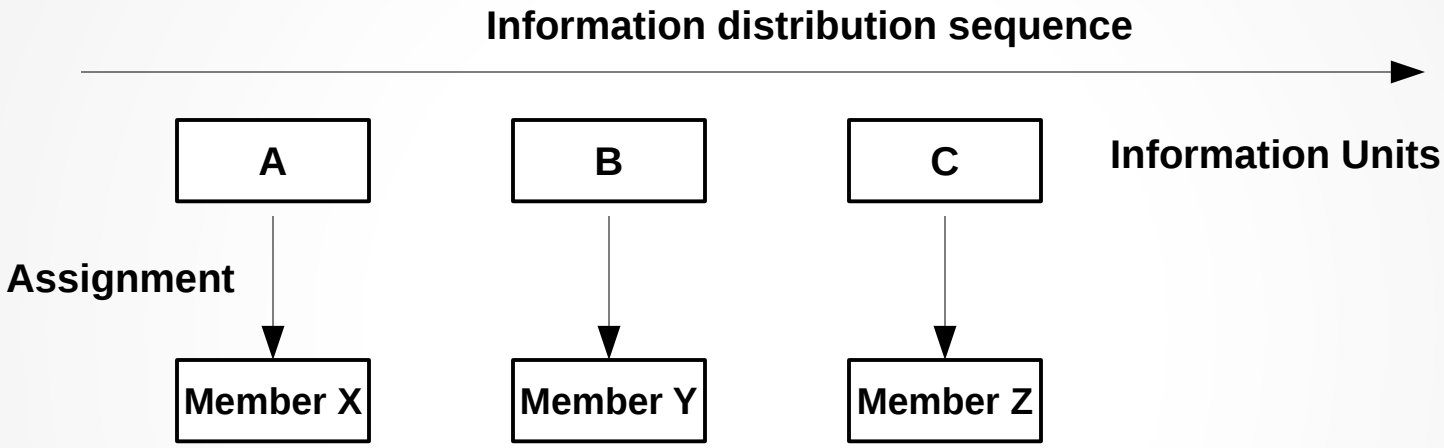
- The network review model is also responsible for industry-wide failure to process required relevant information, making e-discovery meaningless.

A First Inherent Problem

A. Knowledge division problem

When case information is divided among N team members, the members can acquire different information. The members will make inconsistent, conflicting, and even wrong coding decisions because they have different knowledge. See an example followed.

Example: Impacts of Information Division (A-B combination)



Three information units (A, B, C) are assigned respectively to members X, Y and Z. All three members will acquire different information units (A, B, C).

If Information A is essential for understanding information unit B, member Y will make a mistake in coding for B.

For example, if B is an email of John Doe without indicating his role, A is an agreement showing John Doe is an attorney for the client. Member Y will make a mistake in privilege call. There is an unlimited number of A-B interactions.

Exemplar A-B Interactions

Unit A	Unit B
Attorney instructions	An experimental data sheet
Indication of trade secret	Trade secret's technical details
Attorney's view on complaint	A consumer's complaint
Discussion of a retention plan	Work force reduction data
Discuss avoiding patent M	Flaw Chart using Patent M
A study of high cost of recall	Public denial of the problem

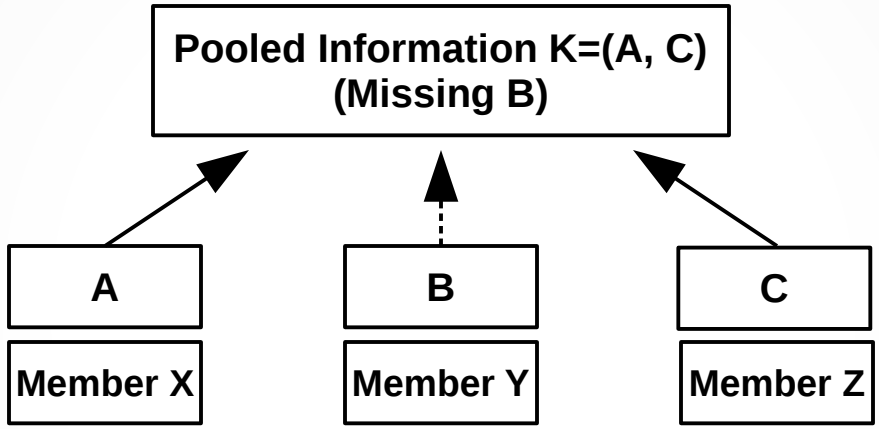
(Some involve three or four unit interactions)

The Second Inherent Problem

B. Difficulty to Pool Essential Information for the Project

A review project is handled by a group of team members, who acquire different information from documents or sources, the failure to collect all member's information will result in an incomplete information pool, and thus directly affects the final determination on the basis of the pooled information.

Example: Impact of Information Deficiency



Three team members X, Y and Z have acquired different information units A, B, and C. All litigation decisions will be made on the basis of pooled information K.

If unit B is essential to the decisions, but is not in the pooled knowledge, the final determination will be impaired. Member's feedback of critical information must be important. If B is unnecessary to the decisions, the omission of B is harmless.

The Third Inherent Problem

C. Complex project requires members' collected knowledge and skills

When a case covers technologies, law, culture, language, and special trade, the required knowledge for properly handling the case is far more than the knowledge of one team member. Thus, there is an inherent need for all term members to share their knowledge to achieve a best result.

Team work is an absolute key to a successful investigation task by law enforcement. It is also true to e-discovery.

Manageable Information Set Concept

The amount of material information is a small fraction of all potential information.

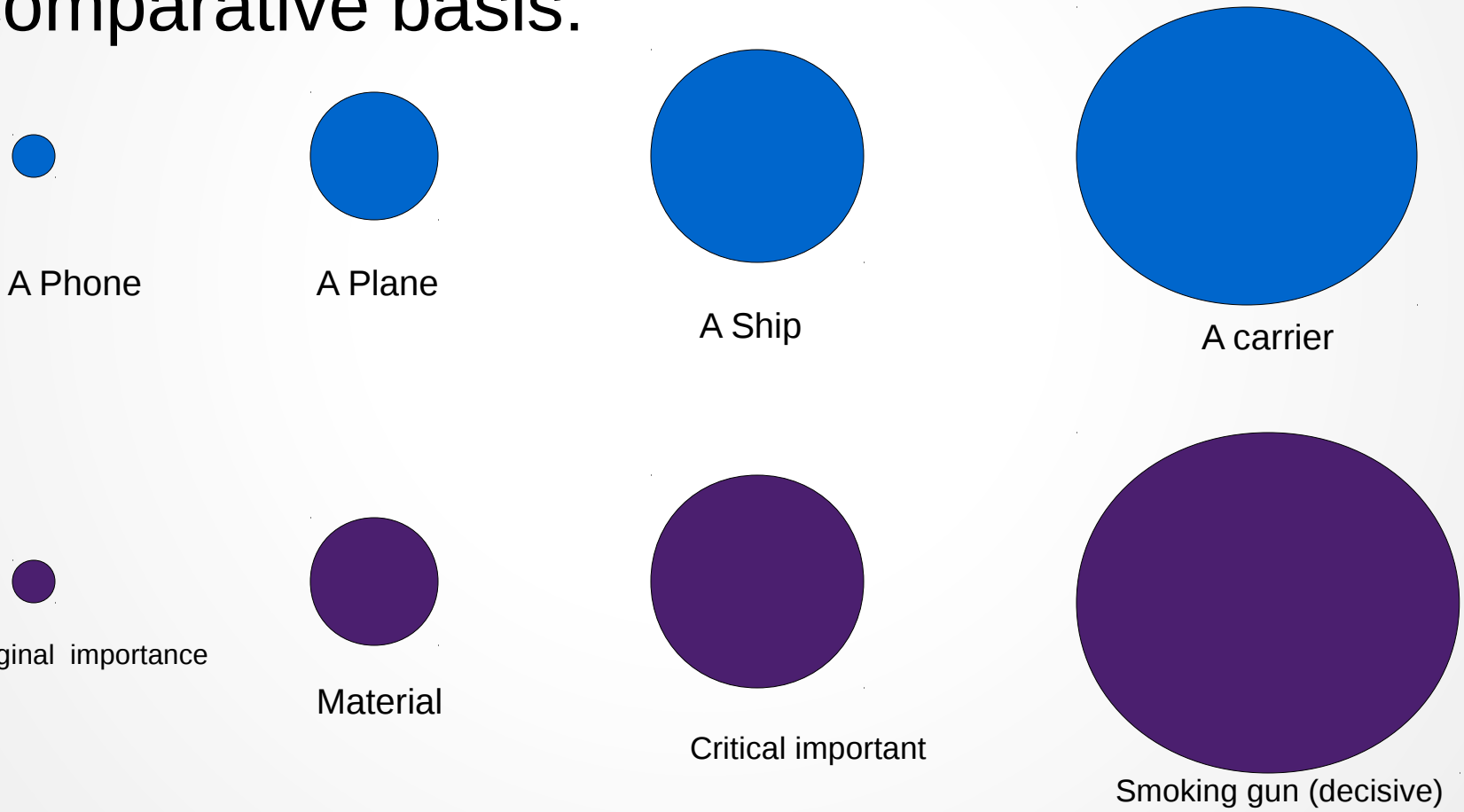
Even though a project may involve a gigantic amount of potential information, the information that actually matters in a case is normally a tiny fraction of the potential volume. Suspects can be quickly narrowed, products may be a few of many, and only a few terms out of millions may be important. So, the amount of material information is only about thousands to ends of thousands of pieces. The amount can be easily handled.

Why the Problems Cannot Be Assessed at Review Site?

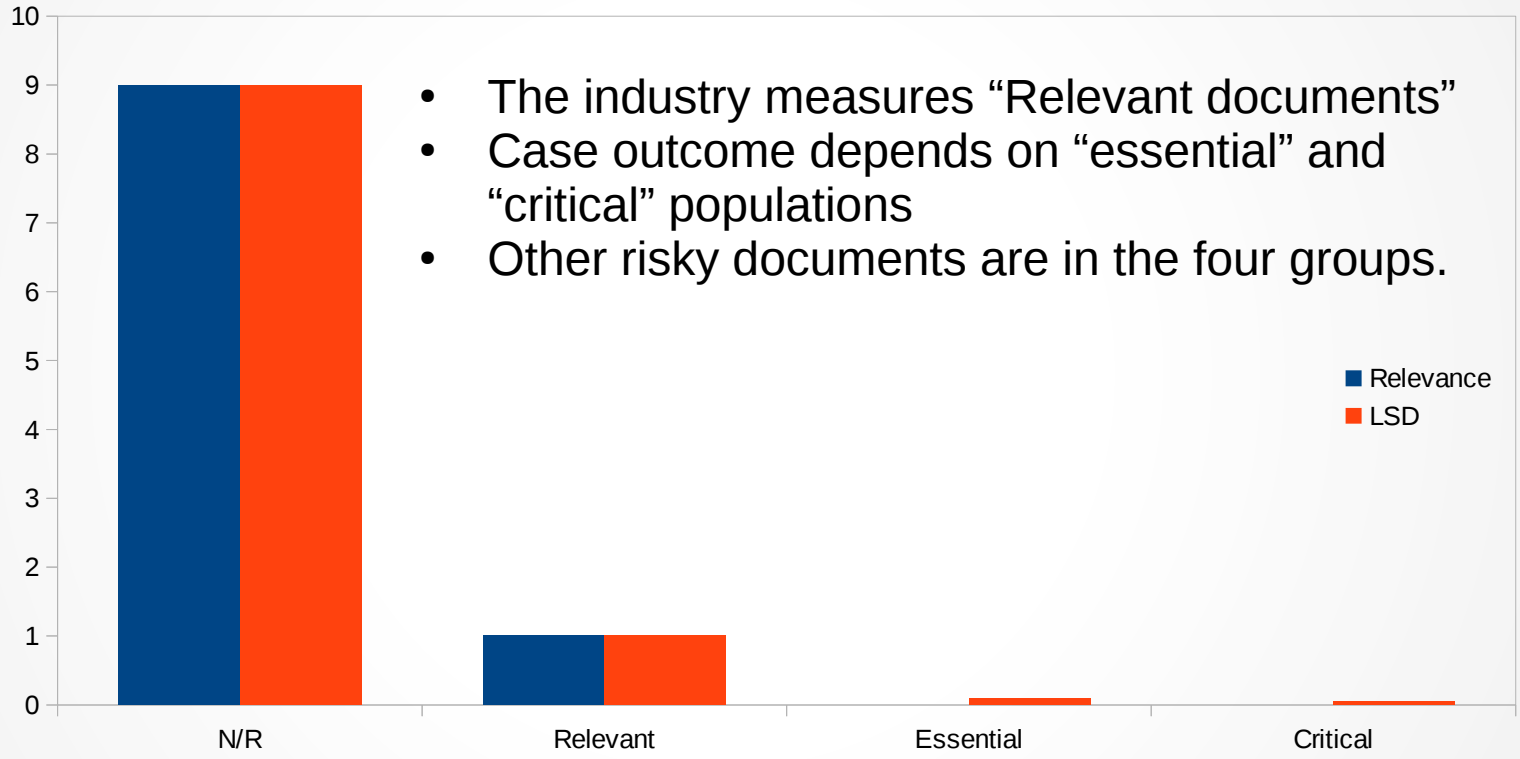
It is because of many flaws in the e-discovery foundation.

Flaw 1: Use Relevance Concept

- Relevance is compared despite it lacks any comparative basis:



Flaw 2: Equating Relevant documents to Ligation Important documents



- The industry measures “Relevant documents”
- Case outcome depends on “essential” and “critical” populations
- Other risky documents are in the four groups.

Flaw 3: Use Error Rates

Error rates are derived from “incorrectly” coded relevant tags, and thus inhere two flaws:

- Relevancy is misused as statistics even though it lacks a common comparative basis.
- When error rates are used, the number count must be important. However, the number count is meaningless because only a small fraction of relevant documents affect case outcome.
- Of course, no court ever decided cases on relevant document number.

Flaw 4: Tasks Are Limited by Cost Proportionality

- So-called proportionality test focuses on case stake and discovery costs only, and thus driving vendors to reduce review costs.
- This equation fails to consider how leaked information damages the company, triggers chain litigation, impairs competitiveness, and loses trade secrets. When a company fails, its impacts will reach stock owners, employees/their families, third-party contractors, and even federal and state governments (re: tax and possible bailout).
- The current test fails to consider individual financial strength.
- Misguided by this principle, corporations ignore duty to control risks for stock owners.

Flaw 5: Misuse Statistical Methods

- Statistical methods are misused for following reasons:
- “Relevancy” is a proper term for defining items or individuals.
- Try to extend relevancy data to review performance that actually depends upon a small number of essential and critical documents.
- Small probability theory is misused to study performance which depends upon extremely low probabilities (e.g., several in a million).
- Most publications (including Peck's ruling on predictive coding) are based upon junk science.

Flaw 6: Failure to Consider Concept-Concept Interactions

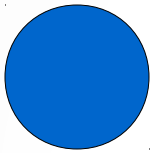
- Most business documents contain only narrow subjects for business convenience. They are not written as standalone documents for judges. Interpretation of them in isolation of related facts can cause reviewers to ignore their relevance and grave risks carried on the documents. See slide 5.

Flaw 7: Allocate Same Review Time

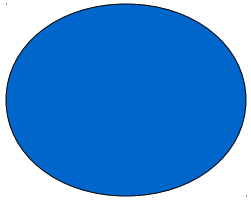
Vendors misuses review quota (doc. no per hour).

A. Production time:

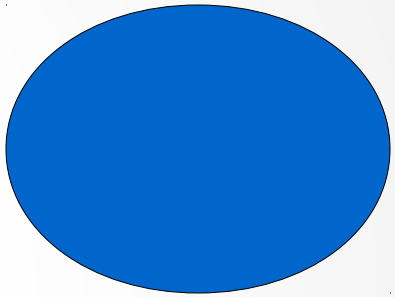
A paper clip



A sofa



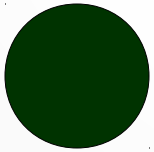
A plane



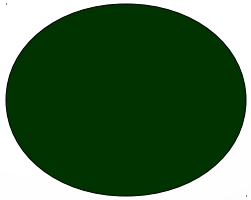
A carrier

B. Review time:

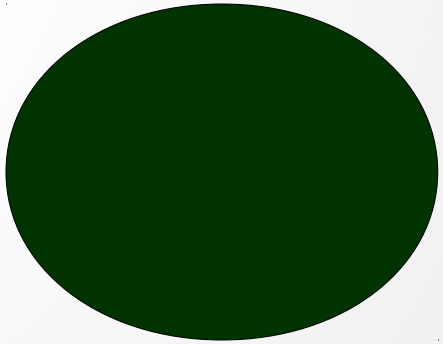
One line email



Short article

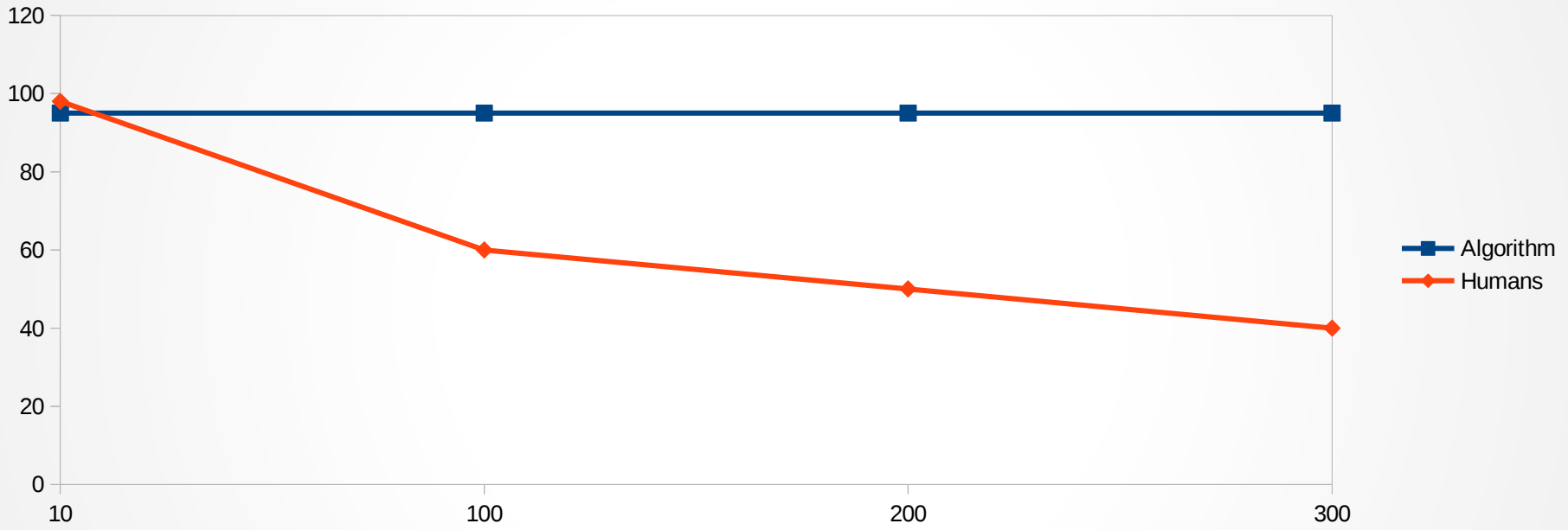


Long article



Compilation or database

Flaw 8: Failure to See Human Review Performance Characteristics



- When a vendor requires high speed review, a computer algorithm defeats humans in coding “accuracy”. Computers are much faster.

Flaw 9: Assume One Correct Tag

- Litigation is like a master game with a large number of rules (case laws and procedural rules). The significance of documents often depends on current claims, potential new claims, and current defenses and potential new defense, as well as what facts are available and what fasts are in dispute.
- It is an improper practice to let a computer algorithm to judge what is right and what is wrong, based upon initial human instructions which is obsolete.

Flaw 10: Failure to See Duplicates' Effects

- In coding two documents A and B, a coding error of A will result in a “50% error rate.”
- If 1000 copies of B is created by a server and a computer is able to get the same answer for all Bs, now, its accuracy is inflated to (1+1000)/(2+1000)=99.9%. The human coding accurate rate will be decreased. If Bs are near duplicates, human accuracy rates would be much lower.
- Thus, “accuracy rate” is absolutely meaningless because backup times could change performance ratings.
- Naturally, critical and sensitive documents are backed with much lower chances.

The Impacts of Meaningless Performance Metrics

- Based upon the foregoing, performance evaluation methods are unable to correctly rate review quality for any of the following reasons:

  Lack of relevancy definition, improper focus on document number, imposing high review speeds, inadequate reviewer incentive, contribution of duplicates, misuse of statistical methods, and incompetent methods. Most ratings are products of junk science, over junk science, over junk science.

- Impacts: When review quality cannot be reasonably rated, vendors can sell any useless and meaningless services to clients. When no body can prove his work is better than others, low prices is the only thing that clients can actually shop for. Therefore, all clients shop for the lowest costs.

Impacts of The e-Discovery Flaws

- The flaws are in the foundation of e-discovery and have interfered with delivery of justice. The following effects must be presumed:
  - Turn cases into games with unpredictable results.
  - Responsible for loss of trade secrets, business secrets, and strategical information.
  - Lead to chain litigation, class actions, and/or damages to litigant company's reputation and good wills.

Exemplar Impacts: Privilege document leak

- *J-M Manufacturing Company, Inc. v. McDermott Will & Emery,* California Superior Court, Los Angeles County – Central District, Case No.: BC 462832 is the first case on privilege leak.
- J-M produced 4000 privileged documents in the first round production. McDermott subsequently re-produced privileged documents 3,900 to the government in the second production. An unidentified entity subsequently produced these documents to the whistleblower.
- Judge Wu noted that “each time J-M learned of this mistaken production of privileged materials, and then re-produced some or all of those same materials, its case for having taken ‘reasonable steps’ would seem to get weaker and weaker."
- Can anyone tell how to prevent such a fatal error? The problem is inherent in the network document review model!

Exemplar impact: loss of trade secrets

- When trade secrets are dispersed on documents, while document reviewers are unable to identify them due to their limited understandings of the subjects, such secrets will be produced in different parts of documents. Those who acquire the documents can figure out the trade secrets.
- Strategical plans and customer information can be lost for the same reason.

Exemplar Impact: Product Liability

A. Inadvertently produced non-relevant information:

A document discusses a prevalent problem of product X and potential improvements.

B. A consumer has following information:

The combination will give the consumer a reasonable ground for filing a product liability lawsuit.

Exemplar impact: Patent Infringement

A. Client produced Non-relevant information:

A document discussed difficult to design around patent M.

B. A Patent holder has following information:

U.S Patent M

Product appear to use Patent M

Willful Infringement?

The combination will give the patent holder

A willful infringement lawsuit for triple damages.

Exemplar impact: Employment liability

A. Client Inadvertently produced following non-relevant information:

A document discussed criteria for terminating some employees.

B. A terminated employee has following information:

Termination

Questionable Firing Pattern

Discriminatory?

The combination will support a discrimination lawsuit. (Event if the motivation does not reach classes, it may hurt the client's reputation). Only one document can cause all damages.

Exemplar impact: Contract Dispute

A. Client Inadvertently produced following non-relevant information:

A product was derived from a compilation N.

B. The compilation N's owner has following information:

Compilation N's Similar Work

Client-owner Contract

Entitle fees per contract?

The combination will support the compilation owner's claim for a share of revenues per their own contract.

Exemplar impacts: Regulatory Action

A. Client Inadvertently produced Non-relevant information:

A product shows the source is from prohibited region.

B. A federal agency has received following information:

Wood Product	Adverse Impact of the product	A legal basis for Issue subpoena?

The agency could not conduct investigation until it heard that wood was harvested in prohibited region.

Damages Propagation

- The above slides show that adverse parties only need one piece of critical document to support a lawsuit. This piece of document often is very innocent when it is viewed standalone.
- When a client does its business, it must have disclosed an overwhelming amount of basic information through products, signed agreements, employee communications, customer services, and public filings (6-K, 8-K, 10-K etc.).
- Future adverse parties need only one or two pieces of information.

How to Address e-Discovery Inherent Problems?

Traditional Tools Are Very Inefficient

Group meetings, phone calls, email distribution, and working together are extremely ineffective.

(1) Potential information scope is so large that one can plan for specific information. Potential suspects, potential terms, potential products etc. are so numerous that it impossible to cover them by traditional methods.

(2) Team members always have different problems, know different information, and need different help. Forcing them to do the same is a massive waste of human labors (>90% wasted time).

Traditional Tools Are Very Inefficient

(3) Gap problem: When a problem is discussed in a meeting, all work done before the meeting cannot be accurately identified, and there is an uncertainty in some documents.

(4) Unworkable: When a large number of issues are discussed on a daily basis, the amount of works that might require corrections cannot be identified for corrective review.

Custom Software Is Unrealistic

- Development of custom database software is impractical due to a very long development cycle.
- Costs for six-phase tasks is PROHIBITORY.
  - Great overhead expenses for designing software;
  - High expense per instruction
- New release is generally full of bugs in the first a few years, but a litigation project is often over in two years.
- Team members have to learn new methods for one-time use. This poses a high risk of use errors.
- Managers have to address unexpected problems at project sites.

FastAction for Solving Problems

- FastAction addresses knowledge deficiency by providing a channel to collect information from team members.
- FastAction cures the knowledge division problem by sharing knowledge in two steps:
  - Sharing project knowledge in real time
  - Revisiting all actions/decisions made before the sharing of each of piece of project knowledge.
  - The scheme is a best solution, but its effectiveness depends upon the efforts of all team members

Why Is FastAction The Best Tool

This is due to low to extremely low usage frequency and the fast database search capability of FastAction.

- In many investigations, document reviews, and researches, information is used in low frequency. Information cannot be collected as inventory “stock”.

- When a team member needs particular information, the member can search for it. FastAction allows for need-based collaboration. This provides a possibility for a team to “collaborate,” seeking hundreds of pieces of useful information from a gigantic information space.

Other Information Available

- FastAction Owner's Use Instructions
- Using FastAction To Remedy Network Review Model Defects
- Using FastAction to Improve Privilege Log
- Using FastAction to Improve Translations For Litigation Purposes
- Using FastAction to Control Corporate Risks.

Contact Information

Jianqing Wu, Ph.D.
A & U Data

tempaddr2@atozpatent.com
202-560-3000