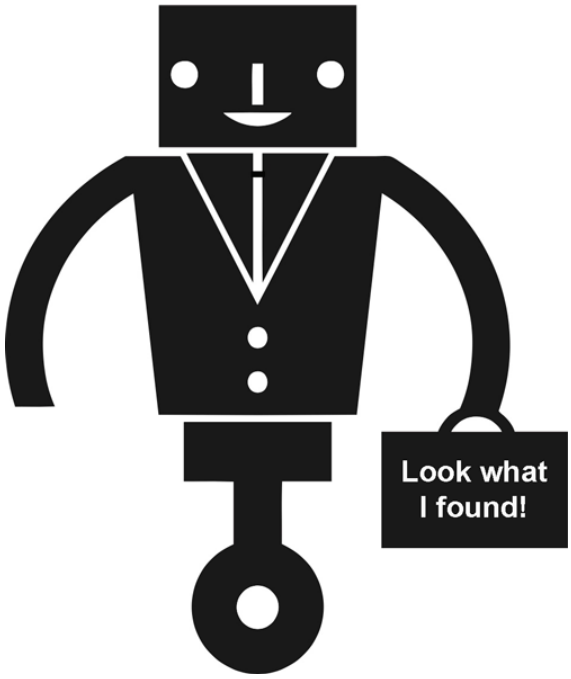
e-Discovery Team

TREC 2016 Total Recall Track

NOTEBOOK

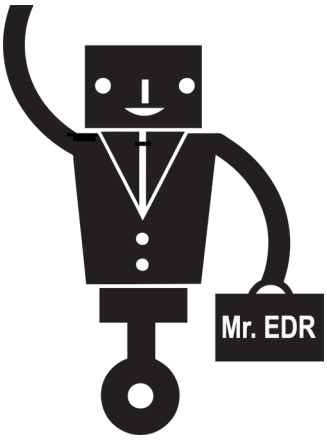
October 25, 2016
Revised December 19, 2016



A collaborative effort of
Ralph Losey, e-DiscoveryTeam.com, e-Discovery Team, LLC
and Kroll Ontrack, Inc., eDiscovery.com.

TABLE OF CONTENTS

e-Discovery Team Members.



The Team is made up of five legal search experts Ralph Losey, Jim Sullivan, Tony Reichenberger, Levi Kuehn, Jani Grantz -- and one "robot," *Mr. EDR* (the software they used). The team members are not scientists or in academia. Most are lawyers who spend their working hours looking for evidence in large, chaotic datasets, such as email. They typically assist other attorneys in lawsuits and legal investigations. Their work includes the identification, review, analysis, classification, production, and admission of Electronically Stored Information (ESI) as evidence in courts in the Unites States and elsewhere.

The Team leader is Ralph C. Losey, J.D., a full-time practicing attorney, principal and *National e-Discovery Counsel* of Jackson Lewis P.C., a U.S. law firm with over 800 attorneys and fifty-five offices. He has over 36 years of experience doing legal document reviews. Losey is also a blogger at *e-DiscoveryTeam.com* where he has written over two million words on e-discovery. He has also written six books published by the American Bar Association and West Thompson. The past five years Losey has participated in multiple public and private experiments, some competitive, to test and prove various predictive coding methods. Losey has also written over sixty articles on the subject of legal search and predictive coding.

Jim Sullivan, J.D., Tony Reichenberger, J.D.,and Jani Grantz J.D., are attorney search and review specialists who work for Kroll Ontrack, Inc. (KO). Levi Kuehn is a non-attorney search and review specialists who works for KO. Kroll Ontrack is the primary e-discovery vendor used by Losey and his law firm. It is a global e-Discovery software, processing and project management company (eDiscovery.com). The Team robot, *Mr. EDR*, is the Team's personalization of Kroll Ontrack's software, *eDiscovery.com Review* (EDR). Losey, Sullivan and Reichenberger participated in the 2015 TREC Total Recall Track. So too did a prior version of Mr. EDR, which is in a process of constant enhancement. The software version used in 2016 contained the latest beta-test version of the software that has not yet been released to the public

Research Questions Considered at TREC 2015 Recall Track

Background to questions considered: It is generally accepted in the legal search community that the use of *predictive coding* type search algorithms can improve the search and review of documents in legal proceedings.[1] The use of predictive coding has also been approved,

[1] *Predictive Coding* is defined by The Grossman-Cormack Glossary of Technology-Assisted Review, 2013 Fed. Cts. L. Rev. 7 (January 2013) (*Grossman-Cormack Glossary*) as: "An industry-specific term generally used to describe a Technology Assisted Review process involving the use of a Machine Learning Algorithm to distinguish Relevant from Non-Relevant Documents, based on

and even encouraged by various courts around the world, including numerous courts in the U.S.[2]

Although there is agreement on use of predictive coding, there is controversy and disagreement as to the most effective ***methods*** of use.[3] There are proponents for a variety of different methods to find training documents for predictive coding. Some advocate for the use of chance selection alone, others for the use of top ranked documents alone, others for a combination of top ranked and mid-level ranked documents where classification is unsure.[4] The-Discovery Team uses a method that includes a combination of all three of these selection processes and more.

Some attorneys and predictive coding software vendors advocate for the use of predictive coding search methods alone, and forego other search methods when they do so, such as keyword search, concept searches, similarity searches and linear review. The e-Discovery Team members reject that approach and instead advocate for a *hybrid multimodal* approach they call *Predictive Coding 4.0*.[5] This method uses an approach to active machine learning that the Team calls ***IST***, standing for "***Intelligently Spaced Training***." Under IST the attorney in charge decides exactly when to train. This is different from other systems where

---

Subject Matter Expert(s) Coding of a Training Set of Documents. " A Technology Assisted Review process is defined as: "A process for Prioritizing or Coding a Collection of electronic Documents using a computerized system that harnesses human judgments of one or more Subject Matter Expert(s) on a smaller set of Documents and then extrapolates those judgments to the remaining Document Collection. … TAR processes generally incorporate Statistical Models and/or Sampling techniques to guide the process and to measure overall system effectiveness." *Also see*: Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, Richmond Journal of Law and Technology, Vol. XVII, Issue 3, Article 11 (2011).

[2] Da Silva Moore v. Publicis Groupe 868 F. Supp. 2d 137 (SDNY 2012) and numerous cases later citing to and following this landmark decision by Judge Andrew Peck, including another more recent opinion by Judge Peck, *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125 (S.D.N.Y. 2015).

[3] Grossman & Cormack, Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery, SIGIR'14, July 6–11, 2014; Grossman & Cormack, Comments on "The Implications of Rule 26(g) on the Use of Technology-Assisted Review", 7 Federal Courts Law Review 286 (2014); Herbert Roitblat, series of five OrcaTec blog posts (1, 2, 3, 4, 5), May-August 2014; Herbert Roitblat, Daubert, Rule 26(g) and the eDiscovery Turkey OrcaTec blog, August 11th, 2014; Hickman & Schieneman, The Implications of Rule 26(g) on the Use of Technology-Assisted Review, 7 FED. CTS. L. REV. 239 (2013); Losey, R. Predictive Coding 3.0, part one (e-Discovery Team 10/11/15).

[4] Id.; Webber, Random vs active selection of training examples in e-discovery (Evaluating e-Discovery blog, 7/14/14).

[5] Losey, R., Predictive Coding 4.0 – Nine Key Points of Legal Document Review and an Updated Statement of Our Workflow (e-Discovery Team, 9/12/16) (Part One of an Eight Part Series explaining the recent advancements from our Predictive Coding method from version 3.0 to version 4.0).

the machine retrains after each document is coded, or certain predetermined number, and the human trainer has no discretion as to timing.[6]

The *e-Discovery Team* approach includes all types of search methods (thus the term *multimodal*) to find relevant documents, with primary reliance placed on predictive coding. The *Team* also uses a variety of methods to find suitable training documents for predictive coding, including high ranking documents, and all other search methods. This is a fundamental difference with other methods that rely entirely on predictive coding to find relevant documents, and rely entirely upon high-ranking documents for training. Grossman and Cormack have scientifically tested these high-ranking training methods, and measured their effectiveness, but this does not mean that they endorse them as an exclusive tool, nor claim this to be their own preferred method.[7]

Four Research Questions:

1. Primary Question (repeat from 2015): What Recall, Precision and Effort levels will the *e-Discovery Team* attain in TREC test conditions over all thirty-four topics using the Team's *Predictive Coding 4.0* hybrid multimodal search methods and Kroll Ontrack's software, *eDiscovery.com Review* (EDR).
2. What is the impact of incorrect Subject Matter Expert ("SME") judgments by the TREC assessors on Recall and Precision. (Unplanned question that unfortunately arose out of the circumstances encountered.)
3. What is the most effective search method from the Team's multimodal tool-set for retrieval of relevant documents in the relatively simplistic search challenges presented by most, but not all, of the thirty-four topics. (Unplanned question that arose out of the circumstances encountered.)
4. What is the role of active machine learning in retrieval of relevant documents in the simplistic search challenges presented by most of the thirty-four topics. (Unplanned question related to the third issue above that also arose out of the circumstances encountered.)

Overview Of Team Participation in 2016 TREC Recall Track

The *e-Discovery Team* participated in all thirty-four of the Total Recall Track *Athome topics*. It did not participate in the fully automated TREC Total Recall sandbox. All thirty-four topics searched a collection of public emails of former Florida Governor Jeb Bush. There were 290,099 emails in the *Jeb Bush Email* collection. In the version of the Jeb Bush emails used by TREC almost all metadata of these emails has been removed. Moreover, the associated

[6] The merits of the Team's approach to the timing of machine learning are detailed in *Predictive Coding 4.0* Part Two.

[7] Grossman & Cormack, *Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery*, SIGIR'14, July 6–11, 2014.

attachments and images were not present. Other collections of the Jeb Bush email exist from PST files that include more information, but the Team did not utilize this information and limited its efforts and attention to the official TREC collection.

This same Jeb Bush email collection was used by the Total Recall Track in 2015 for ten topics. In 2015 Losey searched all ten of these ten topics. None of these search topics was repeated in 2016.

The thirty-four topics searched in 2016, and their names, are shown below. On the far right column are the first names of the e-Discovery Team member who did the review for that topic. The thirteen topics in red were considered mandatory by TREC and the remaining twenty-one were optional. The e-Discovery Team did all topics.

Topic	Name	Reviewer
401	Summer Olympics	Ralph
402	Space	Tony
403	Bottled Water	Ralph
404	Eminent Domain	Tony
405	Newt Gingrich	Ralph
406	Felon Disenfranchisement	Ralph
407	Faith Based Initiatives	Ralph
408	Invasive Species	Tony
409	Climate Change	Levi
410	Condominiums	Tony
411	Stand Your Ground	Ralph
412	2000 Recount	Tony
413	James V. Crosby	Jim
414	Medicaid Reform	Tony
415	George W. Bush	Jim
416	Marketing	Jim
417	Movie Gallery	Ralph
418	War Preparations	Tony
419	Lost Foster Child Rilya Wilson	Levi
420	Billboards	Jim
421	Traffic Cameras	Jim
422	Non Resident Aliens	Tony
423	National Rifle Association	Tony
424	Gulf Drilling	Levi
425	Civil Rights Act of 2003	Ralph
426	Jeffrey Goldhagen	Ralph

427	**Slot Machines**	**Jim**
428	**New Stadiums and Arenas**	**Levi**
429	**Elian Gonzalez**	**Jim**
430	**Restraints and Helmets**	**Jani**
431	**Agency Credit Ratings**	**Tony**
432	**Gay Adoption**	**Jani**
433	**Abstinence**	**Jim**
434	**Bacardi Trademark**	**Ralph**

Ralph Losey did ten topics, Tony Reichenberger did ten, Jim Sullivan did eight, Levi Kuehn did four, and Jani Grantz did two. Unlike the Team's 2015 effort, no contract review attorneys were utilized on any topic. They were all solo efforts, although there was some coordination and commutations between team members on the SME type issues encountered. This pertained to questions of true relevance and errors found in the gold standard for most of these topics.

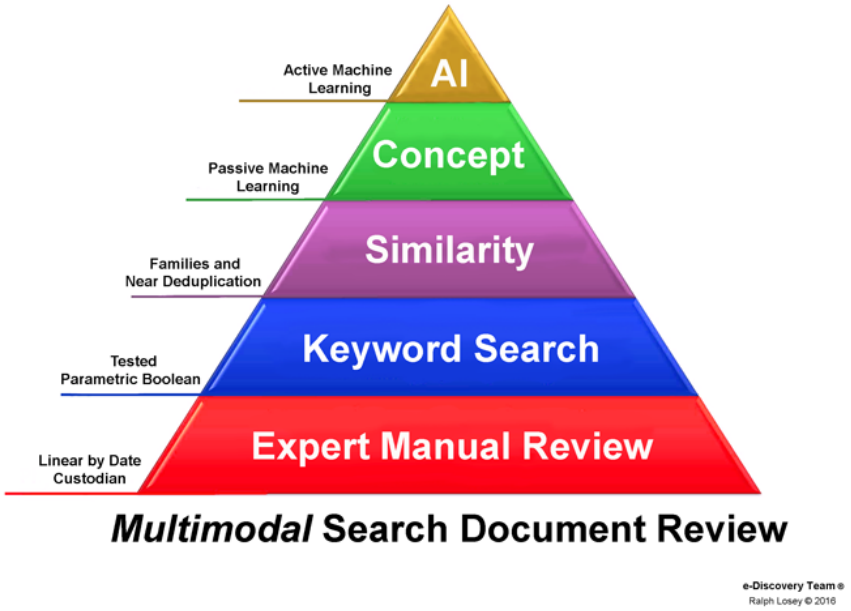
In each Topic the assigned Team attorney personally read and evaluated for true relevance every email that TREC returned as a relevant document, and every email that TREC unexpectedly returned as Irrelevant. Some of these were read and studied multiple times before we made our final calls on true relevance, determinations that took into consideration and gave some deference to the TREC assessor adjudications, but were not bound by them. Many other emails that the Team members considered irrelevant, and TREC agreed, were also personally reviewed as part of their search efforts. As mentioned, there was sometimes consultations and discussion between team members as to the unexpected TREC opinions on relevance.

All of the thirty-four topics presented search challenges to the Team that were easier, some far easier, than the Team typically face as attorneys leading legal document review projects. They were roughly equivalent to the most simplistic challenges that they might face in projects involving very simple legal disputes. A few of the search topics included legal issues, much more than were found in the 2015 Total Recall Track. This is a revision that the Team requested and appreciated because it allowed testing of legal judgment and analysis in determination of true relevance in these topics. In legal search such skills are obviously very important. In most of the 2016 Total Recall topics, however, no special legal training or analysis was required for a determination of true relevance. The Team's final report will specifically identify each topic and, as the Team did in its 2015 TREC report, provide full details on the types of searches performed for each topic and difficulties encountered.

Summary of the Team's Work

The e-Discovery Team's 2016 Total Recall Track *Athome* project started June 3, 2016, and concluded on August 31, 2016. Using a single expert reviewer in each topic the e-Discovery Team classified 9,863,366 documents in 34 different review projects.

All attorneys used the e-Discovery *Team's Predictive Coding 4.0 hybrid multimodal IST* search techniques and were assisted by the KO software, EDR. They relied on active machine learning and other search techniques to find relevant documents and effective training documents. The various types of searches included in the Team's multimodal approach are shown in the search pyramid, below.

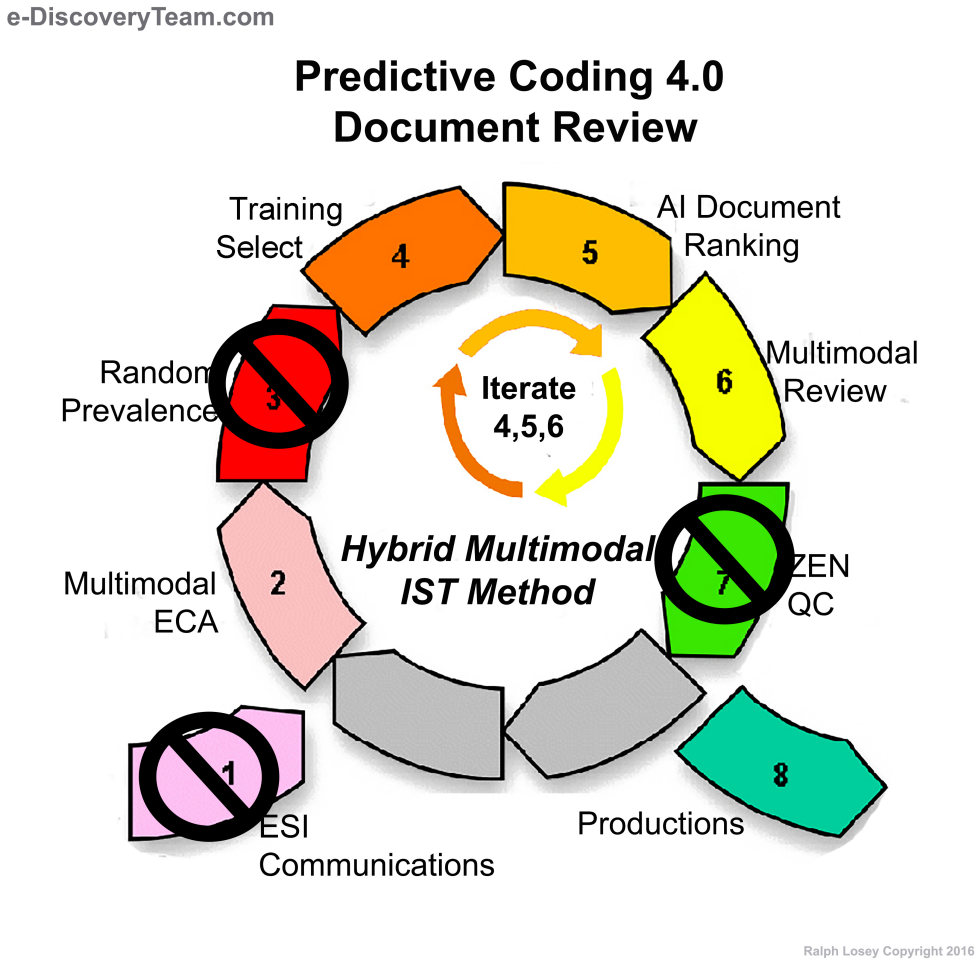


Linear review refers to an SME's examination of all documents by certain key witnesses in a lawsuit during certain time frames critical to the disputed facts in a lawsuit.
Keyword search in our methodology refers to the use of terms originating from legal and document analysis, and from witness interviews. Judgmental sampling and verification by SMEs are also used to test the terms before they are used throughout a document collection. Our keyword search also includes a variety of Boolean functions and parametric targeting, wherein searches are limited to certain metadata fields of an electronic document. Similarity and concept searches refer to a variety of *passive* machine learning analytic search techniques. The AI search at the top of the pyramid refers to the use of active machine learning. The EDR KO software uses a proprietary type of logistic regression algorithm.

The standard eight-step workflow used by the Team in legal search projects is shown in the diagram below.[8] To meet the Team's self imposed time requirements of completing every review project with minimal time efforts, the standard steps Three and Seven were omitted

[8] Losey, R., *Predictive Coding 4.0 – Nine Key Points of Legal Document Review and an Updated Statement of Our Workflow* (e-Discovery Team, October 2016) contains a complete description of all eight steps in parts Six and Seven.

as will be further explained. Further, due to the set-up of the TREC experiments, the first step of our workflow, ESI Communications, was severely constrained to the point of being practically meaningless, as will also be further explained. The Team's standard workflow was thus reduced to five steps as shown below.



In the first step of ESI Communications team members on a legal review project typically spend hours in discussion and analysis of scope of relevance and the target documents. The communications often include hundreds of written exchanges, both informal, such as emails and chats, and formal, such as (1) detailed requests for information contained in court documents such a *subpoenas* or *Request For Production*; (2) input from a qualified SME, who is typically a legal expert with deep knowledge of the factual issues in the case, and thus deep knowledge of what the presiding judge in the legal proceeding will hold to be relevant and discoverable; and, (3) dialogues with the party requesting the production of documents to clarify the search target, and other parties. The ESI communications may lead to formal motions with the governing court, legal memorandums, hearings before the presiding judge and opinions rendered by one or more judges on the scope of relevance.[9]

[9] *Id.* at Part Six wherein the first step of ESI Communications is explained in detail.

The only ESI communications in the TREC experimental set-up was a very short, one sentence description of relevance for each topic. Two topics had a two-sentence description (410-Condominiums and 423-National Rifle Association). The only other type of ESI communications in this TREC Track were the automated, instant returns of all documents submitted as to whether TREC considered them to be relevant or not. There were no appeals or other procedures set-up for *Athome* participants who actually examined the documents for true relevance to challenge obvious errors in judgment.

Short Answers to Research Questions

Research Question # 1 (Primary Question): What Recall, Precision and Effort levels will the *e-Discovery Team* attain in TREC test conditions over all thirty-four Topics using the Team's *Predictive Coding 4.0* hybrid multimodal IST search methods and Kroll Ontrack's software, *eDiscovery Review* (EDR).

Short Answer to Primary Question: Again, like last year, the Team attained excellent results with high levels of Recall and Precision in all topics, including perfect or near perfect results in several topics using the corrected gold standard. The Team was able to do so even though it only used five of the eight steps in its usual methodology, and even though it intentionally severely constrained the amount of human effort expended on each topic. The Team's enthusiasm for the results, which were significantly better than its 2015 effort, is tempered by the fact that the search challenges presented in most of the topics in 2016 were not difficult. As mentioned, they were equivalent to an easy legal search project, such as a simple, single plaintiff employment law dispute. The Final Report will include a detailed analysis of these results.

Research Question # 2: What is the impact of multiple errors in SME judgments by the TREC assessors on Recall and Precision.

Short Answer: The impact on Recall and Precision using the Team's method is significant and, as you would expect, varied determined to the number of errors made by TREC assessors in a particular topic. After the Team encountered numerous errors on the first topics undertaken, it was forced to create its own gold standard of true relevant documents for each topic. The Team's new gold standard corrected for the obvious errors seen in TREC's assessments of relevance. In all close questions on relevance the judgment of TREC's assessors was accepted as accurate.

The obvious errors and inconsistencies seen by the Team's close study of the documents were not accepted. In most, but not all topics, the Team did not use the documents with obvious errors for its machine training. This will be further detailed in the Final Report. In all topics the Team created its own standard and made comparative recall, precision and F1 calculations based thereon. The observation and correction of TREC errors in gold standard became a collaborative effort among the Team to peer review and verify our corrected

standard. Most of these efforts, many of which occurred after the conclusion of the Track in August, were not included in the time reports of efforts expended by attorneys in the search.

The Team was very reluctant to take this step and would certainly have let pass a few errors or mere differences of opinion. We recognize that no standard is ever perfect. As lawyers the Team understands all too well that some, perhaps many judgments on relevance are subjective. Again, in all close questions on relevance the judgments of TREC's assessors were accepted, even though we personally disagreed.

The Team means no disrespect by the creation of an alternate gold standard. We appreciate and respect the efforts made by the TREC assessors and organizers. Still, the volume of obvious errors encountered forced us to take this action. The integrity of our primary research question to test the effectiveness of our hands-on type of *ad hoc* hybrid methods demanded that we do so. We understand that the impact on other *Total Recall* Participants, ones that never actually examine documents, would be far less, perhaps even negligible. Still, there could be an impact, even for them, in some topics where more than an insignificant number of the same or similar documents were inconsistently judged.

The decision to not accept the errors seen, and to instead create our own gold standard, resulted in substantial additional work for the Team. In some topics we even took the step of making two "reasonable calls." One was for TREC, and the second call, which always took place on the next submission, was for our own internal tracking. In the second call we would include emails that we knew from prior submissions of the same or similar document would again be incorrectly considered irrelevant by TREC. We knew they were true relevant and so waited until after our public *reasonable call* to TREC to submit them and then we make our own internal *reasonable call*. We were attempting to, in effect, *play two games at once*, and maximize our score in each game. Keeping track of two standards added an unexpected layer of difficulty to our work and we did not bother to do so in all topics. The *dual-call topics* will be specifically identified in our Final Report.

In some topics the difference between the two standards was substantial. In a few topics it was minor. Some differences were found in all topics. This is not unexpected in any standard involving at least somewhat subjective mass relevance adjudications. We do not intend to engage in a criticism of the specific gold standard creation methods used in 2016 *Total Recall Track*, except to note that the appeals procedure included in the 2008 and 2009 TREC Legal Tracks could have improved the accuracy of the results for the *Total Recall Track Athome* participants.[10] Further, the Team understands from informal reports that the TREC

[10] Participant appeal rights could have mitigated the errors seen in 2016, but this can be burdensome and, as seen in those Tracks in 2008 and 2009, can create their own issues. *See*: Oard, Hedlin, Tomlinson, Baron, *Overview of the TREC 2008 Legal Track,* found at http://trec.nist.gov/pubs/trec17/papers/LEGAL.OVERVIEW08.pdf; and Oard, Hedlin, Tomlinson,

assessors work was much more time constrained than was the work of the Team. Moreover, unlike the Team, the TREC assessors did not have the benefit of SME input from a native Floridian lawyer (Losey) who was familiar with Florida politics and Governor Bush and, since 2015, had put substantial time reviewing this email collection.

The Final Report will include a detailed comparison of recall, precision and F1 based on the comparison of both the TREC and Team assessments. A few examples of the more egregious errors encountered will be provided. The Final Report may also contain a complete listing of the revised gold standards that the Team created for each topic, or at least a conditional offer of disclosure of the corrected standards. The Team invites input from other participants and organizers of the Total Recall Track on this issue. Again, the Team recognizes that no gold standard is ever perfect, including its own revised standards. This will be set forth in further detail in the Team's final report.

Research Question # 3: What is the most effective search method from the Team's multimodal tool-set for retrieval of relevant documents for the relatively simplistic search challenges presented by most of the thirty-four topics.

Short Answer: For the easy topics the Team found that what it calls "tested, parametric, Boolean keyword search" was the most effective search method to find relevant documents. The Team was surprised by how well a sophisticated use of keywords was able to identify nearly all of the target relevant documents in many of the topics in this year's *Total Recall Track*. This shows the continued importance of a multimodal approach to legal search, including especially keyword search, when done properly, especially in simple lawsuits involving relatively easy search issues. This will be set forth in further detail in the Team's final report.

Research Question # 4: What is the role of active machine learning in retrieval of relevant documents in the simplistic search challenges presented by most of the thirty-four topics.

Short Answer: The Team found that for the relatively easy topics in this year's *Total Recall Track* the role of active machine learning was reduced to a quality control function. It would find a few relevant documents not located by keyword search, or concept and similarity search, and thus improve recall somewhat. In the simplest topics active machine learning did not find *any* new relevant documents, but instead only confirmed that all relevant documents had already been found by the other methods. This will be set forth in further detail in the Team's final report.

Further Discussion of Research Question # 1

Baron, Oard, *Overview of the TREC 2009 Legal Track* found at: http://trec.nist.gov/pubs/trec18/papers/LEGAL09.OVERVIEW.pdf.

Even using the given uncorrected TREC standard for scoring, and even though in most topics we did not train on the TREC returned-relevant documents that the Team considered irrelevant, the Team overall still attained excellent results. Under the corrected standard, which will be shared in the Final Report, the results were much better. The following chart compares the Team's Recall, Precision and F-Measure for each *Athome* topic with the results obtained by TREC's BMI and BMI-Desc runs (only other scores now available).

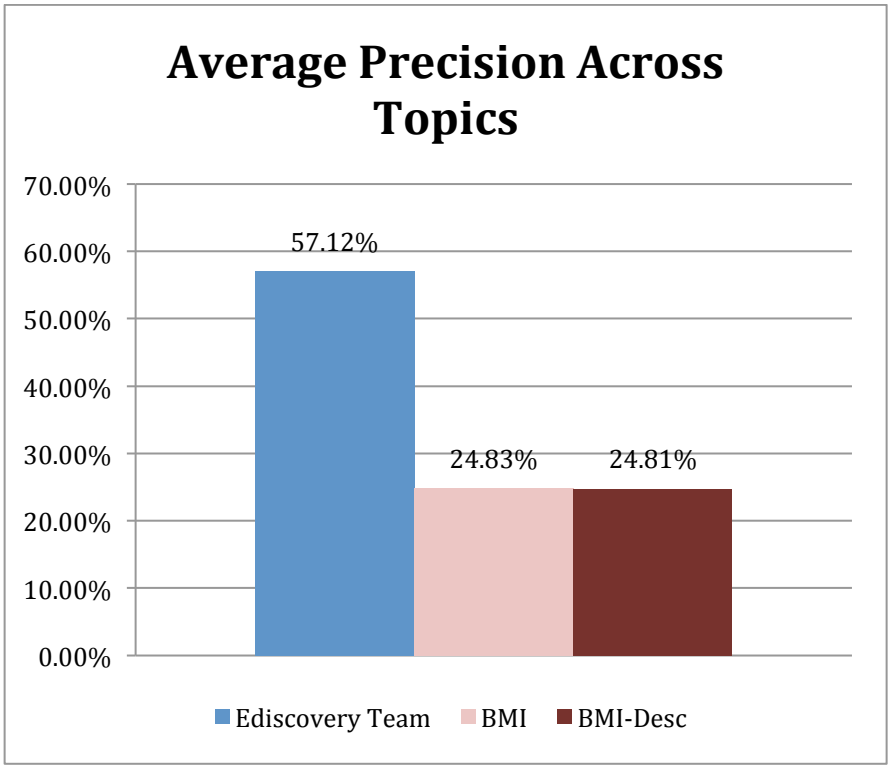
REASONABLE COMPARISON

		Recall			Precision			F-Measure		
		Ediscovery Team	**BMI**	**BMI-Desc**	**Ediscovery Team**	**BMI**	**BMI-Desc**	**Ediscovery Team**	**BMI**	**BMI-Desc**
athome401	**Summer Olympics**	41.05%	91.70%	**92.58%**	**73.44%**	15.31%	15.45%	**52.66%**	26.23%	26.48%
athome402	Space	72.57%	**91.07%**	90.28%	22.04%	**30.86%**	30.59%	33.81%	**46.09%**	45.70%
athome403	Bottled Water	7.16%	**97.71%**	**97.71%**	**80.41%**	37.49%	37.49%	13.14%	**54.18%**	**54.18%**
athome404	Eminent Domain	22.94%	91.74%	**91.93%**	**64.43%**	26.55%	26.61%	33.83%	41.19%	**41.27%**
athome405	**Newt Gingrich**	95.08%	**99.18%**	98.36%	**28.09%**	9.82%	9.74%	**43.36%**	17.87%	17.73%
athome406	**Felon Disenfran**	73.23%	**92.91%**	**92.91%**	**66.91%**	9.58%	9.58%	**69.92%**	17.37%	17.37%
athome407	Faith Based Initiatives	31.02%	91.80%	**91.99%**	**68.72%**	41.86%	41.95%	42.75%	57.50%	**57.62%**
athome408	**Invasive Species**	55.17%	**83.62%**	**83.62%**	**64.65%**	7.87%	7.87%	**59.53%**	14.39%	14.39%
athome409	**Climate Change**	84.65%	**95.05%**	94.06%	**40.71%**	13.99%	13.85%	**54.98%**	24.40%	24.14%
athome410	**Condominiums**	95.10%	**99.48%**	99.03%	**46.13%**	42.59%	42.40%	**62.12%**	59.64%	59.38%
athome411	**Stand Your Ground**	66.29%	70.79%	**84.27%**	**67.05%**	5.70%	6.09%	**66.67%**	10.55%	11.36%
athome412	2000 Recount	57.38%	91.35%	**92.48%**	**49.18%**	40.97%	41.48%	52.96%	**56.57%**	57.27%
athome413	**James V. Crosby**	96.34%	99.08%	**99.27%**	**89.00%**	28.73%	28.78%	**92.52%**	44.55%	44.63%
athome414	Medicaid Reform	91.66%	96.90%	**97.26%**	**35.32%**	35.10%	35.23%	51.01%	51.54%	**51.73%**
athome415	**George W. Bush**	**94.08%**	63.39%	67.08%	**91.04%**	61.09%	58.66%	**92.53%**	62.22%	62.59%
athome416	Marketing	60.30%	94.19%	**95.57%**	42.08%	43.32%	**43.96%**	49.57%	59.35%	**60.22%**
athome417	**Movie Gallery**	99.61%	**99.81%**	99.66%	**99.38%**	57.28%	57.19%	**99.49%**	72.79%	72.67%
athome418	**War Preparations**	39.57%	93.05%	**93.58%**	**50.34%**	12.68%	12.76%	**44.31%**	22.32%	22.45%
athome419	Lost Foster Child Rilya Wilson	**98.84%**	93.06%	93.61%	15.04%	48.13%	**48.41%**	26.10%	63.44%	**63.82%**
athome420	**Billboards**	92.54%	**99.46%**	99.32%	**92.16%**	31.65%	31.61%	**92.35%**	48.02%	47.95%
athome421	**Traffic Cameras**	90.48%	**100.00%**	**100.00%**	**12.50%**	1.90%	1.90%	**21.97%**	3.73%	3.73%
athome422	Non Resident Aliens	93.55%	**100.00%**	**100.00%**	0.90%	**2.81%**	**2.81%**	1.79%	5.46%	5.46%
athome423	**National Rifle Association**	51.05%	**99.65%**	**99.65%**	**33.18%**	18.68%	18.68%	**40.22%**	31.46%	31.46%
athome424	Gulf Drilling	99.60%	**100.00%**	**100.00%**	22.76%	**26.39%**	**26.39%**	37.05%	**41.76%**	**41.76%**
athome425	**CivilRights Act 2003**	91.32%	**98.60%**	**98.60%**	**96.59%**	33.70%	33.70%	**93.88%**	50.23%	50.23%
athome426	**Jeffrey Goldhagen**	70.00%	**94.17%**	**94.17%**	**87.50%**	9.17%	9.17%	**77.78%**	16.72%	16.72%
athome427	**Slot Machines**	89.21%	**96.68%**	**96.68%**	**35.77%**	16.98%	16.98%	**51.07%**	28.89%	28.89%
athome428	New Stadiums	93.10%	**98.49%**	**98.49%**	17.81%	**26.95%**	**26.95%**	29.91%	**42.31%**	**42.31%**
athome429	**Elian Gonzalez**	**94.20%**	99.27%	99.27%	**92.41%**	35.45%	35.45%	**93.29%**	52.24%	52.24%

athome430	**Restraints and Helmets**	71.95%	94.25%	**94.65%**	**65.00%**	36.40%	36.55%	**68.30%**	52.52%	52.74%
athome431	**Agency Credit Rate**	75.69%	**99.31%**	**99.31%**	**47.60%**	11.61%	11.61%	**58.45%**	20.78%	20.78%
athome432	**Gay Adoption**	85.00%	**98.57%**	**98.57%**	**86.23%**	11.20%	11.20%	**85.61%**	20.12%	20.12%
athome433	**Abstinence**	99.11%	**100.00%**	**100.00%**	**66.07%**	9.09%	9.09%	**79.29%**	16.67%	16.67%
athome434	**Bacardi Trademark**	86.84%	**100.00%**	**100.00%**	**91.67%**	3.44%	3.44%	**89.19%**	6.65%	6.65%

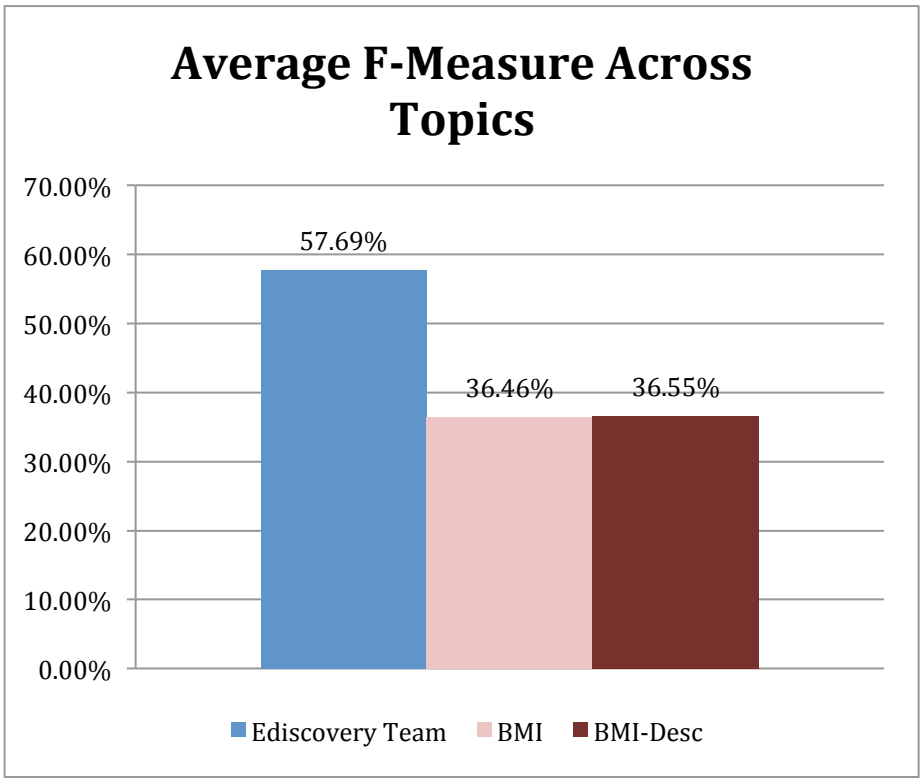
These comparative statistics show the scores at the time of reasonable call.

In the precision category, which in Legal Search is the *money shot* that has the greatest impact on the cost of a document review project, the *e-Discovery Team* dominated. It had the highest precision level on 28 of the 34 topics (82%). They are highlighted in blue in the above chart. The e-Discovery Team's average precision score was 57.1%. The average precision of both BMI and BMI-Desc was 24.8%. Thus the Team's precision score was on average ***more two and a quarter times higher*** than that of the BMI standards.



In the F1-measure, which is the standard value used in legal search to evaluate overall precision and recall of a project, the *e-Discovery Team* again dominated. This is somewhat surprising in view of the fact that these measurements were based on the error-filled TREC standard. The Team had the highest F1 scores on 23 of the 34 topics (68%). They are highlighted in blue in the above chart. The e-Discovery Team's average F1 score was 57.69%.

The average F1 of BMI and BMI-Desc was 36.5%. Thus the Team's F1 score was on average ***more than 58% higher*** *than that of the BMI standards*.



Even using TREC's challenged standard, the Team still attained higher recall than both the BMI and BMI-Desc standards on two topics: topic 415 *George Bush* with a score of **94.08%**; and, topic 419 *Lost Foster Child Rilya Wilson* with a score of **98.84%**. Moreover, the Team attained recall levels in excess of 90% at the time of reasonable call in the following additional topics:

- **95.08%** on topic 406 Felon Disenfranchisement;
- **95.10%** on topic 410 Condominiums;
- **96.34%** on topic 413 James V. Crosby;
- **99.61%** on topic 417 Movie Gallery;
- **92.54%** on topic 420 Billboards;
- **90.48%** on topic 421 Traffic Cameras;
- **93.55%** on topic 422 Non Resident Aliens;
- **99.60%** on topic 424 Gulf Drilling;

- **91.32%** on topic 425 Civil Rights Act of 2003;
- **93.10%** on topic 428 New Stadiums and Arenas;
- **94.20%** on topic 429 Elian Gonzalez;
- **99.11%** on topic 433 Abstinence.

In summary, even with the TREC standard, where in most topics the Team did not use all documents returned as relevant for all of its training documents, it attained Recall scores greater than 90% in fourteen of the thirty-four topics. The Team attained Recall scores of 80% or higher in four additional topics. The average results obtained across all thirty-four topics at the time of reasonable call were as follows:

- 75.46% Recall
- 57.12% Precision
- 57.69% F1
- 121 Docs Reviewed Effort

The Team will disclose all of its scores under the corrected gold standard in the Final Report. In the meantime, here are the average results obtained across all thirty-four topics at the time of reasonable call:

- 87.15% Recall
- 64.94% Precision
- 68.74% F1
- 124 Docs Reviewed Effort

At the time of reasonable call the Team had recall scores greater than 90% in twenty-one of the thirty-four topics and greater than 80% in five more topics. Recall of greater than 99% was attained in seven topics.

At the time of reasonable call the Team had precision scores greater than 90% in thirteen of the thirty-four topics and greater than 80% in two more topics. Precision of greater than 98% was attained in six topics.

At the time of reasonable call the Team had F1 scores greater than 90% in twelve of the thirty-four topics and greater than 80% in one more topic. F1 of greater than 97% was attained in five topics. We were lucky to attain one perfect score, as we did in 2015, in topic (417) with an F1 score of 100%. The perfect score was obtained by locating all 5,945 documents relevant under the corrected standard after reviewing only 45 documents. This topic was filled with form letters and was a fairly simple search. Still, the BMI and BMI-Desc F1 scores for this topic were both under 73%. The Team was pleased to prove, once again, that perfect recall and perfect precision is possible, albeit rare, using the Team's methods.

For questions, comments or suggestions concerning this preliminary Notebook report of the *e-Discovery Team* please contact Ralph.Losey@gmail.com.