

## Trademark/Service Mark Application, Principal Register

**Serial Number: 86634255**  
**Filing Date: 05/19/2015**

**The table below presents the data as entered.**

Input Field	Entered
<b>SERIAL NUMBER</b>	86634255
<b>MARK INFORMATION</b>	
<b>*MARK</b>	<a href="#">CONTINUOUS ACTIVE LEARNING</a>
<b>STANDARD CHARACTERS</b>	YES
<b>USPTO-GENERATED IMAGE</b>	YES
<b>LITERAL ELEMENT</b>	CONTINUOUS ACTIVE LEARNING
<b>MARK STATEMENT</b>	The mark consists of standard characters, without claim to any particular font, style, size, or color.
<b>REGISTER</b>	Principal
<b>APPLICANT INFORMATION</b>	
<b>*OWNER OF MARK</b>	Maura Grossman
<b>INTERNAL ADDRESS</b>	[REDACTED]
<b>*STREET</b>	[REDACTED]
<b>*CITY</b>	New York
<b>*STATE</b> (Required for U.S. applicants)	New York
<b>*COUNTRY</b>	United States
<b>*ZIP/POSTAL CODE</b> (Required for U.S. applicants)	[REDACTED]
<b>LEGAL ENTITY INFORMATION</b>	
<b>TYPE</b>	individual
<b>COUNTRY OF CITIZENSHIP</b>	United States
<b>APPLICANT INFORMATION</b>	
<b>*OWNER OF MARK</b>	Gordon V. Cormack
<b>INTERNAL ADDRESS</b>	[REDACTED]
<b>*STREET</b>	[REDACTED]
<b>*CITY</b>	Waterloo, Ontario
<b>*COUNTRY</b>	Canada
<b>*ZIP/POSTAL CODE</b> (Required for U.S. applicants)	[REDACTED]
<b>LEGAL ENTITY INFORMATION</b>	
<b>TYPE</b>	individual

COUNTRY OF CITIZENSHIP	Canada
<b>GOODS AND/OR SERVICES AND BASIS INFORMATION</b>	
INTERNATIONAL CLASS	009
*IDENTIFICATION	Computer software for reviewing, analyzing, categorizing, prioritizing and storing electronic documents for law firm and corporate departments; data conversion of electronic information; computer software for technology-assisted review of electronic documents; computer software for electronic document and information management; computer software for searching, processing, culling, organizing, filtering, prioritizing, analyzing, reviewing, coding, printing and producing electronic documents in the field of legal discovery
FILING BASIS	SECTION 1(a)
FIRST USE ANYWHERE DATE	At least as early as 04/11/2013
FIRST USE IN COMMERCE DATE	At least as early as 04/11/2013
SPECIMEN FILE NAME(S)	
ORIGINAL PDF FILE	<a href="#">SPE00-156471510-20150519110718347722 . Class 9 - both.pdf</a>
CONVERTED PDF FILE(S) (2 pages)	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0003.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0004.JPG</a>
SPECIMEN DESCRIPTION	screenshots of site from which Applicant's software can be downloaded
INTERNATIONAL CLASS	041
*IDENTIFICATION	Educational services, namely, conducting symposia, seminars, conferences, and workshops in the field of legal discovery and technology-assisted review of electronic documents and development and dissemination of scholarly materials in connection therewith; publication of articles, books, and journals
FILING BASIS	SECTION 1(a)
FIRST USE ANYWHERE DATE	At least as early as 04/11/2013
FIRST USE IN COMMERCE DATE	At least as early as 04/11/2013
SPECIMEN FILE NAME(S)	
ORIGINAL PDF FILE	<a href="#">SPE010-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf</a>
CONVERTED PDF FILE(S) (6 pages)	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0005.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0006.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0007.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0008.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0009.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml1\RFA0010.JPG</a>
SPECIMEN DESCRIPTION	excerpts from seminar presentation
INTERNATIONAL CLASS	042
	Research and development of new technology and theory for others in the field of legal discovery and development and dissemination of scholarly materials in connection therewith; Providing non-downloadable software for reviewing, analyzing, categorizing, prioritizing and storing electronic documents for law firm and corporate departments; data conversion of electronic information; providing non-downloadable software for technology-

<b>*IDENTIFICATION</b>	assisted review of electronic documents; providing non-downloadable software for electronic document and information management; providing non-downloadable computer software for preserving, collecting, processing, culling, filtering, searching, organizing, prioritizing, analyzing, reviewing, coding, printing, producing, and presenting electronic documents in the field of legal discovery
<b>FILING BASIS</b>	SECTION 1(a)
<b>FIRST USE ANYWHERE DATE</b>	At least as early as 04/11/2013
<b>FIRST USE IN COMMERCE DATE</b>	At least as early as 04/11/2013
<b>SPECIMEN FILE NAME(S)</b>	
<b>ORIGINAL PDF FILE</b>	<a href="#">SPE0120-156471510-20150519110718347722 . Class 42 - both.pdf</a>
<b>CONVERTED PDF FILE(S) (1 page)</b>	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0011.JPG</a>
<b>ORIGINAL PDF FILE</b>	<a href="#">SPE0120-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf</a>
<b>CONVERTED PDF FILE(S) (6 pages)</b>	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0012.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0013.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0014.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0015.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0016.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0017.JPG</a>
<b>SPECIMEN DESCRIPTION</b>	excerpts from seminar discussing research; screenshot of non-downloadable software
<b>INTERNATIONAL CLASS</b>	045
<b>*IDENTIFICATION</b>	Litigation support services; Legal services, namely, providing customized information through the review of documents for investigation or litigation; Litigation support services, namely, conducting technology-assisted review of electronic documents and electronic legal discovery in the nature of preserving, collecting, processing, culling, filtering, searching, organizing, prioritizing, analyzing, reviewing, coding, printing, producing, and presenting e-mails and other electronically stored information that could be relevant in a lawsuit or government investigation
<b>FILING BASIS</b>	SECTION 1(a)
<b>FIRST USE ANYWHERE DATE</b>	At least as early as 04/11/2013
<b>FIRST USE IN COMMERCE DATE</b>	At least as early as 04/11/2013
<b>SPECIMEN FILE NAME(S)</b>	
<b>ORIGINAL PDF FILE</b>	<a href="#">SPE01230-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf</a>
<b>CONVERTED PDF FILE(S) (6 pages)</b>	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0018.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0019.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0020.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0021.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0022.JPG</a>
	<a href="#">\\TICRS\EXPORT16\IMAGEOUT16\866\342\86634255\xml\RF0023.JPG</a>

<b>SPECIMEN DESCRIPTION</b>	excerpts from seminar presentation regarding litigation support services
<b>ATTORNEY INFORMATION</b>	
<b>NAME</b>	Patricia L. Werner
<b>ATTORNEY DOCKET NUMBER</b>	0338023
<b>FIRM NAME</b>	Bryan Cave LLP
<b>STREET</b>	1290 Avenue of the Americas
<b>CITY</b>	New York
<b>STATE</b>	New York
<b>COUNTRY</b>	United States
<b>ZIP/POSTAL CODE</b>	10104
<b>PHONE</b>	212-541-2341
<b>FAX</b>	212-541-4630
<b>EMAIL ADDRESS</b>	nyuspto@bryancave.com
<b>AUTHORIZED TO COMMUNICATE VIA EMAIL</b>	Yes
<b>OTHER APPOINTED ATTORNEY</b>	Erik W. Kahn, Hassan Albakri, Liz Goldberg, Bonnie Lee, and all other Bryan Cave LLP attorneys
<b>CORRESPONDENCE INFORMATION</b>	
<b>NAME</b>	Patricia L. Werner
<b>FIRM NAME</b>	Bryan Cave LLP
<b>STREET</b>	1290 Avenue of the Americas
<b>CITY</b>	New York
<b>STATE</b>	New York
<b>COUNTRY</b>	United States
<b>ZIP/POSTAL CODE</b>	10104
<b>PHONE</b>	212-541-2341
<b>FAX</b>	212-541-4630
<b>*EMAIL ADDRESS</b>	nyuspto@bryancave.com
<b>*AUTHORIZED TO COMMUNICATE VIA EMAIL</b>	Yes
<b>FEE INFORMATION</b>	
<b>APPLICATION FILING OPTION</b>	TEAS RF
<b>NUMBER OF CLASSES</b>	4
<b>FEE PER CLASS</b>	275
<b>*TOTAL FEE DUE</b>	1100
<b>*TOTAL FEE PAID</b>	1100
<b>SIGNATURE INFORMATION</b>	
<b>SIGNATURE</b>	/patricialwerner/
<b>SIGNATORY'S NAME</b>	Patricia L. Werner
<b>SIGNATORY'S POSITION</b>	Attorney of Record, NY State Bar Member

<b>SIGNATORY'S PHONE NUMBER</b>	212-541-2341
<b>DATE SIGNED</b>	05/19/2015
<b>SIGNATURE</b>	/patricialwerner/
<b>SIGNATORY'S NAME</b>	Patricia L. Werner
<b>SIGNATORY'S POSITION</b>	Attorney of Record, NY State Bar Member
<b>SIGNATORY'S PHONE NUMBER</b>	212-541-2341
<b>DATE SIGNED</b>	05/19/2015

---

## Trademark/Service Mark Application, Principal Register

**Serial Number: 86634255**

**Filing Date: 05/19/2015**

### To the Commissioner for Trademarks:

**MARK:** CONTINUOUS ACTIVE LEARNING (Standard Characters, see [mark](#))

The literal element of the mark consists of CONTINUOUS ACTIVE LEARNING.

The mark consists of standard characters, without claim to any particular font, style, size, or color.

The applicants, Maura Grossman, a citizen of United States, having an address of

[REDACTED]  
New York, New York 10075  
United States

Gordon V. Cormack, a citizen of Canada, having an address of

[REDACTED]  
Waterloo, Ontario N2L5C8  
Canada

request registration of the trademark/service mark identified above in the United States Patent and Trademark Office on the Principal Register established by the Act of July 5, 1946 (15 U.S.C. Section 1051 et seq.), as amended, for the following:

International Class 009: Computer software for reviewing, analyzing, categorizing, prioritizing and storing electronic documents for law firm and corporate departments; data conversion of electronic information; computer software for technology-assisted review of electronic documents; computer software for electronic document and information management; computer software for searching, processing, culling, organizing, filtering, prioritizing, analyzing, reviewing, coding, printing and producing electronic documents in the field of legal discovery

In International Class 009, the mark was first used by the applicant or the applicant's related company or licensee or predecessor in interest at least as early as 04/11/2013, and first used in commerce at least as early as 04/11/2013, and is now in use in such commerce. The applicants are submitting one(or more) specimen(s) showing the mark as used in commerce on or in connection with any item in the class of listed goods/services, consisting of a(n) screenshots of site from which Applicant's software can be downloaded.

**Original PDF file:**

[SPE00-156471510-20150519110718347722 . Class 9 - both.pdf](#)

**Converted PDF file(s)** (2 pages)

[Specimen File1](#)

[Specimen File2](#)

International Class 041: Educational services, namely, conducting symposia, seminars, conferences, and workshops in the field of legal discovery and technology-assisted review of electronic documents and development and dissemination of scholarly materials in connection therewith; publication of articles, books, and journals

In International Class 041, the mark was first used by the applicant or the applicant's related company or licensee or predecessor in interest at least as early as 04/11/2013, and first used in commerce at least as early as 04/11/2013, and is now in use in such commerce. The applicants are submitting one(or more) specimen(s) showing the mark as used in commerce on or in connection with any item in the class of listed goods/services, consisting of a(n) excerpts from seminar presentation.

**Original PDF file:**

[SPE010-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf](#)

**Converted PDF file(s)** (6 pages)

[Specimen File1](#)

[Specimen File2](#)

[Specimen File3](#)  
[Specimen File4](#)  
[Specimen File5](#)  
[Specimen File6](#)

International Class 042: Research and development of new technology and theory for others in the field of legal discovery and development and dissemination of scholarly materials in connection therewith; Providing non-downloadable software for reviewing, analyzing, categorizing, prioritizing and storing electronic documents for law firm and corporate departments; data conversion of electronic information; providing non-downloadable software for technology-assisted review of electronic documents; providing non-downloadable software for electronic document and information management; providing non-downloadable computer software for preserving, collecting, processing, culling, filtering, searching, organizing, prioritizing, analyzing, reviewing, coding, printing, producing, and presenting electronic documents in the field of legal discovery

In International Class 042, the mark was first used by the applicant or the applicant's related company or licensee or predecessor in interest at least as early as 04/11/2013, and first used in commerce at least as early as 04/11/2013, and is now in use in such commerce. The applicants are submitting one(or more) specimen(s) showing the mark as used in commerce on or in connection with any item in the class of listed goods/services, consisting of a(n) excerpts from seminar discussing research; screenshot of non-downloadable software.

**Original PDF file:**

[SPE0120-156471510-20150519110718347722 . Class 42 - both.pdf](#)

**Converted PDF file(s)** (1 page)

[Specimen File1](#)

**Original PDF file:**

[SPE0120-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf](#)

**Converted PDF file(s)** (6 pages)

[Specimen File1](#)

[Specimen File2](#)

[Specimen File3](#)

[Specimen File4](#)

[Specimen File5](#)

[Specimen File6](#)

International Class 045: Litigation support services; Legal services, namely, providing customized information through the review of documents for investigation or litigation; Litigation support services, namely, conducting technology-assisted review of electronic documents and electronic legal discovery in the nature of preserving, collecting, processing, culling, filtering, searching, organizing, prioritizing, analyzing, reviewing, coding, printing, producing, and presenting e-mails and other electronically stored information that could be relevant in a lawsuit or government investigation

In International Class 045, the mark was first used by the applicant or the applicant's related company or licensee or predecessor in interest at least as early as 04/11/2013, and first used in commerce at least as early as 04/11/2013, and is now in use in such commerce. The applicants are submitting one(or more) specimen(s) showing the mark as used in commerce on or in connection with any item in the class of listed goods/services, consisting of a(n) excerpts from seminar presentation regarding litigation support services.

**Original PDF file:**

[SPE01230-156471510-20150519110718347722 . 41 45 - BOTH 2 .pdf](#)

**Converted PDF file(s)** (6 pages)

[Specimen File1](#)

[Specimen File2](#)

[Specimen File3](#)

[Specimen File4](#)

[Specimen File5](#)

[Specimen File6](#)

The applicant's current Attorney Information:

Patricia L. Werner and Erik W. Kahn, Hassan Albakri, Liz Goldberg, Bonnie Lee, and all other Bryan Cave LLP attorneys of Bryan Cave LLP

1290 Avenue of the Americas

New York, New York 10104

United States

The attorney docket/reference number is 0338023.

The applicant's current Correspondence Information:

Patricia L. Werner  
Bryan Cave LLP  
1290 Avenue of the Americas  
New York, New York 10104  
212-541-2341(phone)  
212-541-4630(fax)  
nyuspto@bryancave.com (authorized)

**E-mail Authorization:** I authorize the USPTO to send e-mail correspondence concerning the application to the applicant or applicant's attorney at the e-mail address provided above. I understand that a valid e-mail address must be maintained and that the applicant or the applicant's attorney must file the relevant subsequent application-related submissions via the Trademark Electronic Application System (TEAS). Failure to do so will result in an additional processing fee of \$50 per international class of goods/services.

A fee payment in the amount of \$1100 has been submitted with the application, representing payment for 4 class(es).

#### **Declaration**

The signatory believes that: if the applicant is filing the application under 15 U.S.C. Section 1051(a), the applicant is the owner of the trademark/service mark sought to be registered; the applicant or the applicant's related company or licensee is using the mark in commerce on or in connection with the goods/services in the application, and such use by the applicant's related company or licensee inures to the benefit of the applicant; the specimen(s) shows the mark as used on or in connection with the goods/services in the application; and/or if the applicant filed an application under 15 U.S.C. Section 1051(b), Section 1126(d), and/or Section 1126(e), the applicant is entitled to use the mark in commerce; the applicant has a bona fide intention to use or use through the applicant's related company or licensee the mark in commerce on or in connection with the goods/services in the application. The signatory believes that to the best of the signatory's knowledge and belief, no other person has the right to use the mark in commerce, either in the identical form or in such near resemblance as to be likely, when used on or in connection with the goods/services of such other person, to cause confusion or mistake, or to deceive. The signatory being warned that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. Section 1001, and that such willful false statements and the like may jeopardize the validity of the application or any registration resulting therefrom, declares that all statements made of his/her own knowledge are true and all statements made on information and belief are believed to be true.

#### **Declaration Signature**

Signature: /patricialwerner/ Date: 05/19/2015  
Signatory's Name: Patricia L. Werner  
Signatory's Position: Attorney of Record, NY State Bar Member  
Signature: /patricialwerner/ Date: 05/19/2015  
Signatory's Name: Patricia L. Werner  
Signatory's Position: Attorney of Record, NY State Bar Member  
RAM Sale Number: 86634255  
RAM Accounting Date: 05/19/2015

Serial Number: 86634255  
Internet Transmission Date: Tue May 19 11:22:55 EDT 2015  
TEAS Stamp: USPTO/BAS-XXX.XX.XX.XX-20150519112255619  
186-86634255-530e3db59a870c6f6ccfaf2b14f  
613e06dd8f5be9e145543352867f1ea4f5425c-D  
A-10202-20150519110718347722



# CONTINUOUS ACTIVE LEARNING

# Baseline Model Implementation for Automatic Participation in the TREC 2015 Total Recall Track

The Baseline Model Implementation ("BMI") is free software (licence: [GPL v3](#)) that participants may use to automate the [TREC 2015 Total Recall Track](#). BMI uses [Continuous Active Learning \(CAL\)](#) to fully automate the task required of Total Recall Track participants. The goal for participants is to achieve better effectiveness than BMI, either by modifying it, or by implementing their own solutions from scratch.

## Overview of BMI

To use BMI, participants must:

- Download and install free VirtualBox Software from Oracle.
- Download and install free BMI virtual machine and run scripts.
- [optionally] configure the BMI run scripts.
- [optionally] modify the BMI virtual machine.
- Run the BMI virtual machine using VirtualBox.
- Track progress using logs and/or the Web.
- Repeat until desired results are achieved.

## Installing VirtualBox

Installers for Linux, Mac OS X, and Windows are [available from Oracle](#).

Debian and Debian-based Linux distributions have an available package "virtualbox" that may be installed using a package manager or the command line "sudo apt-get install virtualbox".

Be aware that VirtualBox will run the 64-bit BMI virtual machine only if the BIOS "Virtualization Technology" and "VT-d" features are enabled. It appears that many desktop-class machines disable these features by default. If you don't enable them, BMI will install but fail to boot.

BMI requires 2GB of RAM to run and may consume up to 100GB of disk space for full runs. The host computer on which you install VirtualBox should have substantially more than this.

## Installing BMI

On Linux, ensure that either wget or curl is installed, open a terminal, and type one of these commands:

- `curl plg.uwaterloo.ca/~gvcormac/trecvm/vminstall1.sh | bash`
- `wget -O - plg.uwaterloo.ca/~gvcormac/trecvm/vminstall1.sh | bash`

On Mac OS X, open a terminal and type the following command:

- `curl plg.uwaterloo.ca/~gvcormac/trecvm/vminstall.sh | bash`

On Windows download and run the installer at

- [plg.uwaterloo.ca/~gvcormac/trecvm/vminstall.bat](#)

On all systems, the installer will create a virtual machine named TRECTR, and will install a folder "vmscripts" in the VirtualBox home folder. The name of the VirtualBox home folder is reported by the installer, and can also be determined using VirtualBox.

## Configuring BMI

### Configuring BMI Parameters

In the folder "vmscripts" you will find file "TREC\_Config.txt" which you may edit to supply your group name, the name of the test you wish to run, and the name of the TREC Total Recall server. The default contents of "vmscripts" is:

```
TRUSER=test
TRSERVER=quaid.uwaterloo.ca:33333
TRTEST=trivial
TRRUN=testrun
```

To perform simple testing, you do not need to change any of these settings.

When you receive a participant ID from TREC, you should replace the value of TRUSER with that ID. (The "test" ID is restricted to smaller tests.)

Barring unforeseen circumstances, it should never be necessary to change TRSERVER. TRSERVER is the name of the TREC server from which BMI fetches the datasets, topics, and relevance assessments. The interface to the TREC server [is documented here](#).

TRTEST may (at the time of writing) be one of: "trivial", "test", and "bigtest".

- "trivial" uses 7 topics with 2 datasets containing 30 documents each, and is useful to see that things are working. On typical hardware, "trivial" takes several minutes to run.
- "test" uses 7 topics with 2 datasets containing about 20,000 documents each, and is therefore more useful for evaluating retrieval effectiveness. On typical hardware, "test" takes about 30 minutes.
- "bigtest" uses 2 topics with one dataset containing about 750,000 documents. On typical hardware, "bigtest" takes about 5 hours.

## Configuring BMI Scripts

The folder "vmscripts/cmd.dir" is shared with the BMI virtual machine. Immediately after the BMI VM boots, it executes "vmscripts/cmd.dir/start" which is a bash script. This script can execute any command installed on the BMI VM (which is Debian Linux with developer tools installed). The script can also execute other scripts or (Debian) executable files contained in vmscripts.

The BMI VM is configured with a 2GB root drive and a 1TB scratch drive mounted as /tmp. All drives are reinitialized every time the BMI VM is rebooted.

Log files and other persistent information may be written to files in the vmscripts folder.

The scripts run as the user named "user".

## Configuring the BMI VM

The password for "user" is "user" and the password for "root" is "root". You can sign on to the VM while it is running. However, since the disks are reinitialized at every boot, you will not be able to permanently install software. You can compile and/or install software into the shared folder.

Participants who need to modify the operating system (or to use an entirely different operating system, such as Windows) will need to create a new root disk (possibly by using VirtualBox's "clonehd" to copy the BMI root disk). However, the shared folder and scratch configurations should be preserved, and the VM must read and honor the contents of the TREC\_Config.txt file in the shared folder.

# **Technology-Assisted Review Best Practices: Tools, Techniques, and Validation Methods**

---

**Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz**

**Gordon V. Cormack  
University of Waterloo**

**Arlington, VA  
January 15, 2014**

**WACHTELL, LIPTON, ROSEN & KATZ**

# Today's Agenda Part I: Fundamentals of TAR

---

- The Document Review Problem
  - Traditional Approaches: Search Terms and Manual Review
  - New Approaches: Technology-Assisted Review ("TAR")
- What TAR Is
  - Different Types of TAR and How They Work
- Scientific Evidence for TAR
- Open TAR Issues

# Defining “Technology-Assisted Review” (“TAR”)

---

- A process for **prioritizing** or **categorizing** an entire collection of documents using **computer technologies** that harness **human judgments** of one or more subject matter expert(s) **on a small subset** of the documents, and then **extrapolates those judgments** to the remaining documents in the collection.

Source: Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola*, U.S. Magistrate Judge, 7 Fed. Courts L. Rev. 1 (Jan. 2013), <http://www.fclr.org/fclr/articles/html/2010/grossman.pdf>.

# A High-Level Overview of How Machine Learning Works

---

- **Step 1 – Some documents are selected for review**
  - What documents, how they are selected, and how many documents differ by tool and approach
- **Step 2 – Someone familiar with the case manually reviews those documents**
  - Who should review them is presently a matter of disagreement
- **Step 3 – The TAR tool reviews the rest**

## Variant #4: Continuous Active Learning (“CAL”)

---

- A small, initial set of training documents (a “seed set”) is identified through judgmental sampling.
- A machine learning algorithm is used to identify the most-likely responsive documents in the collection that are not yet included in the “training set.”
- These documents are then reviewed and coded by a human, and added to the training set.
- This process is repeated until the most-likely responsive documents, when reviewed, are found to be non-responsive.
- **There is no separate “review set,”** as the training set contains all of the most-likely responsive documents, and these documents have already been reviewed in the course of CAL.



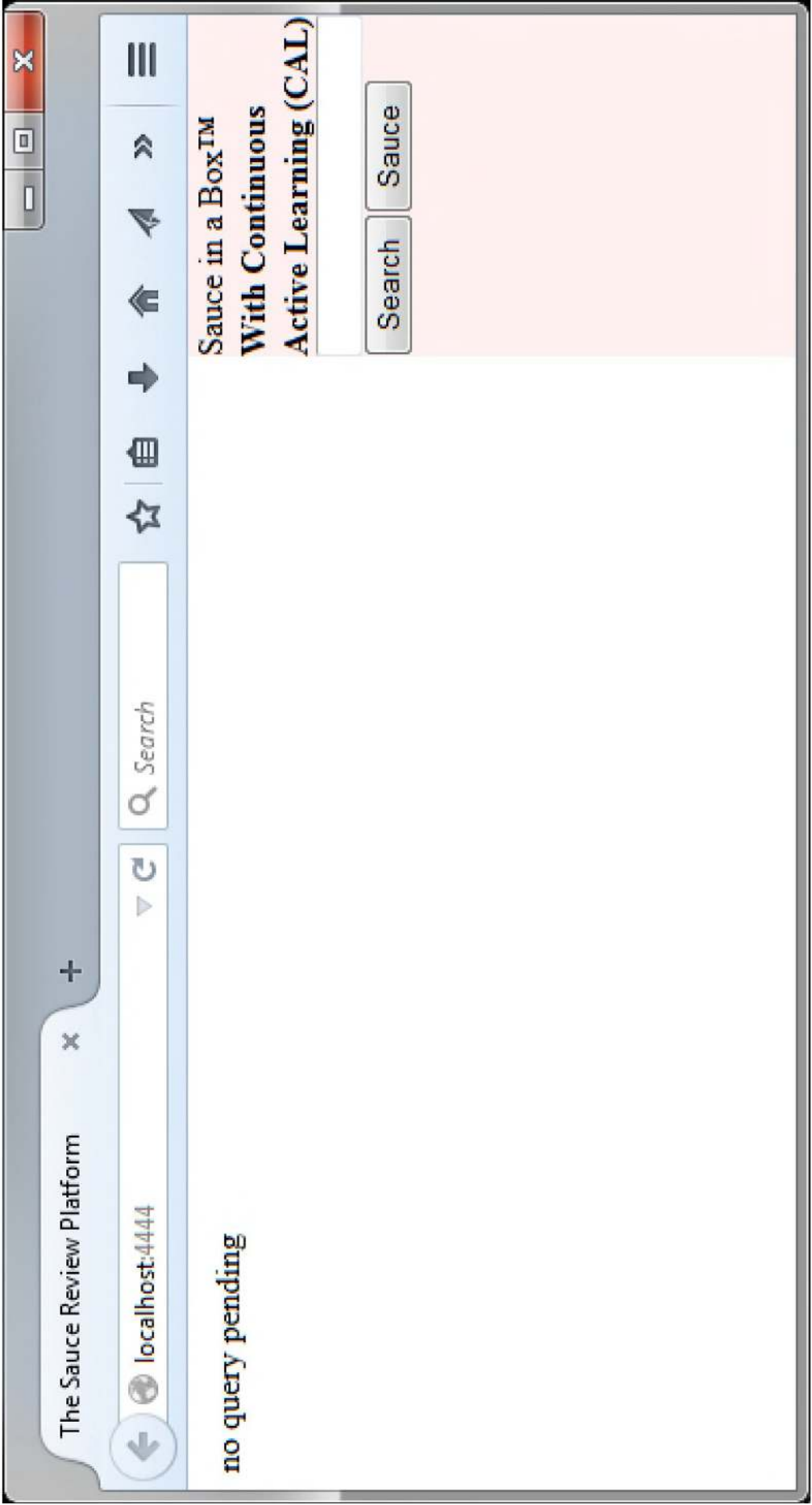
# Contact Information

---

Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz  
[mrgrossman@wlrk.com](mailto:mrgrossman@wlrk.com)  
212-403-1391

Gordon V. Cormack  
University of Waterloo  
[gvcormac@uwaterloo.ca](mailto:gvcormac@uwaterloo.ca)  
519-888-4567 x 34450

WACHTELL, LIPTON, ROSEN & KATZ



# **Technology-Assisted Review Best Practices: Tools, Techniques, and Validation Methods**

---

**Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz**

**Gordon V. Cormack  
University of Waterloo**

**Arlington, VA  
January 15, 2014**

**WACHTELL, LIPTON, ROSEN & KATZ**

# Today's Agenda Part I: Fundamentals of TAR

---

- The Document Review Problem
  - Traditional Approaches: Search Terms and Manual Review
  - New Approaches: Technology-Assisted Review ("TAR")
- What TAR Is
  - Different Types of TAR and How They Work
- Scientific Evidence for TAR
- Open TAR Issues

# Defining “Technology-Assisted Review” (“TAR”)

---

- A process for **prioritizing** or **categorizing** an entire collection of documents using **computer technologies** that harness **human judgments** of one or more subject matter expert(s) **on a small subset** of the documents, and then **extrapolates those judgments** to the remaining documents in the collection.

Source: Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola*, U.S. Magistrate Judge, 7 Fed. Courts L. Rev. 1 (Jan. 2013), <http://www.fclr.org/fclr/articles/html/2010/grossman.pdf>.

# A High-Level Overview of How Machine Learning Works

---

- **Step 1 – Some documents are selected for review**
  - What documents, how they are selected, and how many documents differ by tool and approach
- **Step 2 – Someone familiar with the case manually reviews those documents**
  - Who should review them is presently a matter of disagreement
- **Step 3 – The TAR tool reviews the rest**

# Variant #4: Continuous Active Learning (“CAL”)

---

- A small, initial set of training documents (a “seed set”) is identified through judgmental sampling.
- A machine learning algorithm is used to identify the most-likely responsive documents in the collection that are not yet included in the “training set.”
- These documents are then reviewed and coded by a human, and added to the training set.
- This process is repeated until the most-likely responsive documents, when reviewed, are found to be non-responsive.
- **There is no separate “review set,”** as the training set contains all of the most-likely responsive documents, and these documents have already been reviewed in the course of CAL.

# Contact Information

---

Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz  
[mrgrossman@wlrk.com](mailto:mrgrossman@wlrk.com)  
212-403-1391

Gordon V. Cormack  
University of Waterloo  
[gvcormac@uwaterloo.ca](mailto:gvcormac@uwaterloo.ca)  
519-888-4567 x 34450

WACHTELL, LIPTON, ROSEN & KATZ



# **Technology-Assisted Review Best Practices: Tools, Techniques, and Validation Methods**

---

**Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz**

**Gordon V. Cormack  
University of Waterloo**

**Arlington, VA  
January 15, 2014**

**WACHTELL, LIPTON, ROSEN & KATZ**

# Today's Agenda Part I: Fundamentals of TAR

---

- The Document Review Problem
  - Traditional Approaches: Search Terms and Manual Review
  - New Approaches: Technology-Assisted Review ("TAR")
- What TAR Is
  - Different Types of TAR and How They Work
- Scientific Evidence for TAR
- Open TAR Issues

# Defining “Technology-Assisted Review” (“TAR”)

---

- A process for **prioritizing** or **categorizing** an entire collection of documents using **computer technologies** that harness **human judgments** of one or more subject matter expert(s) **on a small subset** of the documents, and then **extrapolates those judgments** to the remaining documents in the collection.

Source: Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola*, U.S. Magistrate Judge, 7 Fed. Courts L. Rev. 1 (Jan. 2013), <http://www.fclr.org/fclr/articles/html/2010/grossman.pdf>.

# A High-Level Overview of How Machine Learning Works

---

- **Step 1 – Some documents are selected for review**
  - What documents, how they are selected, and how many documents differ by tool and approach
- **Step 2 – Someone familiar with the case manually reviews those documents**
  - Who should review them is presently a matter of disagreement
- **Step 3 – The TAR tool reviews the rest**

## Variant #4: Continuous Active Learning (“CAL”)

---

- A small, initial set of training documents (a “seed set”) is identified through judgmental sampling.
- A machine learning algorithm is used to identify the most-likely responsive documents in the collection that are not yet included in the “training set.”
- These documents are then reviewed and coded by a human, and added to the training set.
- This process is repeated until the most-likely responsive documents, when reviewed, are found to be non-responsive.
- **There is no separate “review set,”** as the training set contains all of the most-likely responsive documents, and these documents have already been reviewed in the course of CAL.

# Contact Information

---

Maura R. Grossman  
Wachtell, Lipton, Rosen & Katz  
[mrgrossman@wlrk.com](mailto:mrgrossman@wlrk.com)  
212-403-1391

Gordon V. Cormack  
University of Waterloo  
[gvcormac@uwaterloo.ca](mailto:gvcormac@uwaterloo.ca)  
519-888-4567 x 34450

WACHTELL, LIPTON, ROSEN & KATZ