

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION**

**FELLOWSHIP FILTERING
TECHNOLOGIES, LLC,**

Plaintiff,

v.

CLOUDERA INC.,

Defendant.

Civil Action No. _____

JURY TRIAL DEMANDED

COMPLAINT FOR PATENT INFRINGEMENT

Plaintiff Fellowship Filtering Technologies, LLC (“Fellowship Filtering” or “Plaintiff”), by and through its attorneys, brings this action and makes the following allegations of patent infringement relating to U.S. Patent No. 5,884,282 (“the ‘282 patent”). Defendant Cloudera, Inc. (“Cloudera” or “Defendant”) infringes Fellowship Filtering’s ‘282 patent in violation of the patent laws of the United States of America, 35 U.S.C. § 1 *et seq.*

INTRODUCTION

1. In a relentless effort to expand its product base and profit from the sale of infringing computer-based recommendation technologies, Cloudera has undertaken to copy the technologies and inventions of Gary Robinson, the inventor or the ‘282 patent and a co-owner of Fellowship Filtering.

2. Cloudera has recognized the importance of providing accurate recommendations using collaborative filtering techniques such as those in the ‘282 patent.

Due to the explosion of web traffic and users the scale of recommendation poses new challenges for recommendation systems. These systems face the dual challenge of producing high quality recommendations and calculating these personalized recommendations for millions of users. In recent years, collaborative filtering (CF) has become a popular way to effectively meet these challenges. CF techniques start

by analyzing the user-item matrix to identify relationships between different users or items, and use that information to produce recommendations for each user.¹

3. Mr. Robinson is a mathematician and inventor of computer-based recommendation engine technologies that enable the recommending of products and/or content based on novel algorithms that calculate the preferences based on the similarity and dissimilarity of users of a website.

4. Mr. Robinson studied mathematics at Bard College and New York University's Courant Institute of Mathematical Sciences. Mr. Robinson is the recipient of the National Science Foundation – SBIR award.

5. Mr. Robinson is a named inventor of numerous United States Patents. Mr. Robinson's patents have been acquired by companies including Google, Inc. ("Google").² Patents referencing Mr. Robinson's '282 patent have been purchased or assigned to companies including: International Business Machines Corporation ("IBM"),³ Google,⁴ Amazon.com, Inc. ("Amazon"),⁵ and Intel Corporation ("Intel").⁶

ROBINSON'S LANDMARK ELECTRONIC MAIL INVENTIONS

6. The Robinson Method, named after Gary Robinson, is a Bayesian statistical approach that uses a text-classifier, rule-based method for determining the relevancy of an email message. Numerous leading SPAM filtering technologies utilize the Robinson Method.⁷

¹ Josh Patterson, *Recommendation with Apache Mahout in CDH3*, CLOUDERA BLOG, November 30, 2011, <http://blog.cloudera.com/blog/2011/11/recommendation-with-apache-mahout-in-cdh3/>.

² See USPTO Assignment Abstract of Title Database Reel/Frame No. 021552/0256.

³ U.S. Patent Nos. 6,356,879; 6,931,397; 7,006,990; 7,080,064; 7,099,859; 7,389,285; 7,885,962; 8,700,448; and 8,825,681.

⁴ U.S. Patent Nos. 7,966,632; 8,290,964; and 8,762,394.

⁵ U.S. Patent Nos. 6,266,649; 7,113,917; 7,433,832; 7,478,054; 7,664,669; 7,778,890; 7,908,183; 7,921,042; 7,945,475; 8,001,003; 8,024,222; 8,108,255; 8,140,391; and 8,180,689.

⁶ U.S. Patent Nos. 6,405,034, 7,590,415, and 7,797,343.

⁷ Ricardo Villamarín-Salomón & José Carlos Brustoloni, *Bayesian Bot Detection Based on DNS Traffic Similarity*, in SAC'09: ACM SYMPOSIUM ON APPLIED COMPUTING 2040—41 (2009); Masahiro Uemura & Toshihiro Tabata, *Design and Evaluation of a Bayesian-filter-based Image Spam Filtering Method*, in PROCEEDINGS OF THE 2008 INTERNATIONAL CONFERENCE ON INFORMATION SECURITY AND ASSURANCE 46-51 (2008) ("the Robinson Method"); MARCO ANTONIO BARRENO, Technical Report No. UCB/EECS-2008-63, EVALUATING THE SECURITY OF MACHINE LEARNING ALGORITHMS 45 (2008); Manabu Iwanaga et al., *Evaluation of Anti-Spam*

7. Mr. Robinson's contributions to the field of electronic mail filtering are recognized as landmark technologies.

Robinson Fisher Method: With the Robinson Fisher method, Gary Robinson developed a more sophisticated way to ensure sensitivity for both recommendations and rejections. Consequently, the Robinson Fisher approach replaced the Geometric Means proposal. To formulate two null hypotheses one must assume ideal conditions, i.e. that token frequencies are pairwise independent, not uniformly distributed, and that the description consists of a random set of tokens. We then calculate a score

GÜNTHER HÖBLING, *PERSONALIZED MEANS OF INTERACTING WITH MULTIMEDIA CONTENT* 119 (2011).

8. Mr. Robinson has published academic articles on statistical approaches to identifying content. A 2003 article in *Linux Journal* described these mathematical approaches for identifying unsolicited bulk email. Mr. Robinson's approach was notable because it assigned scores to both "spam" and "ham" and used an algorithm to guess intelligently whether an incoming email was spam. This approach was incorporated in products such as SpamAssassin, which used a Bayesian statistical approach using a text-classifier rule to distinguish "spam" and "ham" messages.⁸

9. Mr. Robinson's inventions relating to filtering technologies have been widely adopted by spam filters including Spam Assassin⁹ (PC Magazine's Editor's Choice for spam

Methods Combining Bayesian Filtering and Strong Challenge and Response, in PROCEEDINGS OF CNIS'03 (COMMUNICATION, NETWORK, AND INFORMATION SECURITY) 214—19 (2003); BLAINE NELSON, Technical Report No. UCB-EECS-2010-140, BEHAVIOR OF MACHINE LEARNING ALGORITHMS IN ADVERSARIAL ENVIRONMENTS 62-67 (2010); Gordon V. Cormack & Mona Mojdeh, *Autonomous Personal Filtering Improves Global Spam Filter Performance*, in PROCEEDINGS OF THE 6TH CONFERENCE ON EMAIL AND ANTI-SPAM 2 (2009).

⁸ Gary Robinson, *A Statistical Approach to the Spam Problem*, LINUX JOURNAL 107 (2003).

⁹ *SpamAssassin Pro*, in PC MAGAZINE February 25, 2003 at 82 (awarding SpamAssassin Pro its editors' choice award); *The SpamAssassin Project: Train SpamAssassin's Bayesian Classifier*, <http://spamassassin.apache.org/full/3.2.x/doc/sa-learn.html> ("Gary Robinson's f(x) and combining algorithms, as used in SpamAssassin"); *Credits - The Perl Programming Language - Algorithms*, <http://cpansearch.perl.org/src/JMASON/Mail-SpamAssassin-3.2.5/CREDITS> ("The Bayesian-style text classifier used by SpamAssassin's BAYES rules is based on an approach outlined by Gary Robinson. Thanks, Gary!").

filtering), SpamSieve¹⁰ (MacWorld's Software of the Year), and SpamBayes¹¹ (PC Worlds Editor's Choice for spam filtering).

ROBINSON'S DEVELOPMENT OF CONTENT FILTERING SYSTEMS

10. Prior to developing groundbreaking electronic mail filtering technologies, Mr. Robinson used his insights to develop the automated content filtering technologies that are used today by Cloudera and many of the world's largest corporations without attribution or compensation.

11. In the late 1980's, Mr. Robinson developed a system for collecting preference information and providing recommendations. His company, 212-ROMANCE, was an automated, voice-based dating service that used a passive data collection process to determine likely romantic matches.¹² Mr. Robinson's contributions to the field of content filtering were pioneering.



Matthew French, *Romantic Beginnings Have Worldwide Effect*, BOSTON BUS. J., May 20, 2002.

¹⁰ David Progue, *From the Deck of David Progue: The Follow-Up Edition*, N.Y. TIMES, April 5, 2006, <http://www.nytimes.com/2006/04/05/technology/06POGUE-EMAIL.html> ("Spam Sieve is just incredibly, amazingly accurate; my in box is clean, baby, clean!").

¹¹ Tom Spring, *Spam Slayer: 2003 Spam Awards*, PCWORLD MAGAZINE, December 15, 2003, at 36 ("What makes the program unique is that SpamBayes doesn't use predetermined spam definitions. Rather, it constantly evolves by scanning your in-box to build custom definitions."); MARCO ANTONIO BARRENO, Technical Report No. UCB/EECS-2008-63, EVALUATING THE SECURITY OF MACHINE LEARNING ALGORITHMS 45 (2008) ("SpamBayes classifies using token scores based on a simple model of spam status proposed by Robinson . . . SpamBayes Tokenizes the header and body of each email before constructing token spam scores. Robinson's method assumes that each token's presence or absence in an email affects that email's spam status independently from other tokens.").

¹² 212-Romance was incorporated under the name Microvox Systems, Inc.

12. In the mid-1990s, Mr. Robinson recognized that the growing adoption of the internet and increased computational power enabled collection and processing of data relating to customer and user preferences that, with proper data analytics processes, could provide accurate recommendations of products and content.

13. Mr. Robinson further recognized that the growth of the internet led to unique problems involving information overload that filtering techniques using specific new collaborative filtering technologies could solve.

14. At the time, existing recommendation technologies, discussed in the '282 patent, failed to teach a robust and accurate process for providing recommendations. A key insight of Mr. Robinson was that the input of buying habits and/or ratings information from multiple users over the internet allowed similarity values among users to be calculated based on identifying subgroups of similar users.

15. Mr. Robinson invented an automated collaborative filtering ("ACF") system that received and stored data based on internet users' purchasing history, preferences, and/or buying history. When a new user accessed the ACF system through a website (in one embodiment), the ACF system recommended further content (*e.g.*, products) based on the similarity values for the first user as compared with other users that previously provided preference data to the ACF system.

16. Mr. Robinson worked to develop novel systems and processes designed to provide accurate content and product recommendations using data stored, collected, and computed on specific computer-based systems. Mr. Robinson's insights led to the patent application resulting in the '282 patent.

17. The patent-in-suit - the '282 patent - is a pioneering patent in the field of data analytics. The '282 patent uses novel algorithmic approaches to provide accurate recommendations of products and content using data analysis specific to a computer system.

good. The creative license for statistical filtering really belongs to hackers like Paul Graham, Gary Robinson, and Bill Yerazunis and the rest of the community that has invented many of these approaches. Some companies have claimed the technology as their own, which gives people the idea that any other solutions are nonstandard, when it's really borrowed technology.

JONATHAN A. ZDZIARSKI, ENDING SPAM: BAYESIAN CONTENT FILTERING AND THE ART OF STATISTICAL LANGUAGE CLASSIFICATION 269 (2005).

18. The '282 patent has been cited by over 267 United States patents as prior art before the United States Patent and Trademark Office.¹³ Companies whose patents cite the '282 patent include:

- OpenText S.A.
- Accenture Global Services GMBH
- YellowPages.com LLC
- Nielsen Holdings N.V.
- International Business Machines Corporation
- Koninklijke Philips N.V.
- Google, Inc.
- Amazon.com, Inc.
- Microsoft Technology Licensing LLC
- Arbor Networks, Inc.
- Johnson & Johnson Consumer Companies
- S.C. Johnson & Son Inc.
- Sony Electronics, Inc.
- Infosys Ltd.
- Parasoft Corporation
- AT&T Intellectual Property LLP
- Dish Network LLC
- eBay, Inc.
- Rovi Corporation
- CBS Interactive, Inc.
- American Express Company
- Hewlett-Packard Company
- Xerox Corp.
- Capital One Financial Corporation
- JDA Software Group, Inc.
- State University of New York

¹³ The 267 forward citations to the '282 patent do not include patent applications published by the United States Patent and Trademark Office or patent applications that were abandoned prior to publication in the face of the '282 patent.

- Robert Bosch Healthcare System, Inc.
- Netflix, Inc.
- Intel Corporation
- Tribune Media Company
- Ingenio, LLC
- Recommend, Inc.
- Dassault Systemes S.A.
- Pandora Media, Inc.
- Pace plc
- Regents of the University of California
- Facebook, Inc.
- Numera, Inc.

19. Patents citing Mr. Robinson's '282 patent as prior art have been asserted by Amazon.com, Inc. ("Amazon") and Netflix, Inc. ("Netflix") in patent infringement cases:

- Amazon asserted U.S. Patent No. 6,266,649, entitled "Collaborative Recommendations Using Item-to-Item Similarity Mappings," against Discovery Communications, Inc. ("Discovery"). The '649 patent claimed a priority date of September 1998 (subsequent to the '282 patent). Amazon's '649 patent cited Mr. Robinson's '282 patent as prior art during prosecution before the Patent and Trademark Office. After two years of litigation, Discovery took a license to Amazon's '649 patent (prior to claim construction being adjudicated).¹⁴
- Netflix asserted U.S. Patent No. 7,024,381, claiming a priority date of April 2000, against Blockbuster LLC ("Blockbuster"). The '381 patent referenced the '282 patent as prior art. A settlement and license agreement was reached between Netflix and Blockbuster on the verge of trial.¹⁵
- Robert Bosch Healthcare Systems, Inc. ("Robert Bosch") asserted U.S. Patent Nos. 7,223,235 & 7,223,236 against MedApps, Inc. ("MedApps"). The '235 and '236 patents cite Mr. Robinson's '282 patent as prior art. MedApps reached a settlement and license with Robert Bosch roughly one year after the infringement action was initiated.¹⁶
- Black Hills Media LLC ("Black Hills") asserted U.S. Patent Nos. 8,028,323, 8,230,099, and 8,458,356. The '323, '099, and '356 patents referenced Mr. Robinson's '282 patent as prior art. Black Hills settled a majority of its cases following denial of summary judgment of invalidity.¹⁷

¹⁴ *Amazon.com Inc v. Discovery Communications Inc.*, Case No. 09-cv-00681 Dkt. Nos. 122 & 166 (W.D. Wash.).

¹⁵ *Netflix, Inc. v. Blockbuster, Inc.*, Case No. 06-cv-02361 Dkt. No. 239 (Cal. N.D.).

¹⁶ *Robert Bosch Healthcare Systems, Inc. -v- MedApps, Inc.* Case No. 12-cv-00113 Dkt. No. 64 (Cal. N.D.); US. Patent No. 8,028,323 Information Disclosure Statement (March 3, 2010).

¹⁷ *Black Hills Media LLC v. Sonos, Inc.*, Case No. 14-cv-00486 Dkt. Nos. 129 & 169 (Cal. C.D.).

- i2 Technologies, Inc. (“i2”) asserted U.S. Patent No. 7,370,009 against Oracle in the Eastern District of Texas. Subsequently, Oracle asserted four patents against i2’s parent, JDA Software Group. Following a year of litigation, the parties reached a settlement in March 2011.¹⁸

20. Cases against Oracle, Discovery and Blockbuster underscore the inventive nature of the ‘282 patent, as the above asserted cases involve patents referencing Mr. Robinson’s ‘282 patent as prior art.

21. The claims in the ‘282 patent are directed at solving a problem that did not arise in prior art systems, *i.e.* generating preference data from large data sets. In prior art systems, the sample size of users was typically very small, and thus the need for a process that takes into account unusual similarities was not at issue. There is no question pre-electronic recommendation systems are significantly different from computer and/or internet-based recommendation systems. The speed, quantity, and variety of rating information markedly differ from the objectives and data available to recommendation systems existing before modern, computer and/or internet-based systems. Differences between the analog versions of preference systems and the invention disclosed in the ‘282 patent diverge significantly.

22. The use of ratings data and probability values to make recommendations over a computer network was not a longstanding or fundamental economic practice at the time of the invention disclosed in the ‘282 patent. Nor at the time was the use of ratings data and probability values to make recommendations a fundamental principle in ubiquitous use on the internet or computers in general. Dr. Zeynep Tufekci of Harvard University’s Berkman Center for Internet and Society described recommendation engine systems such as the systems disclosed in the ‘282 patent as being far from a “law of nature.”

The fear I have is that every time this is talked about, people talk about it as if it's math or physics, therefore some natural, neutral world. And they're programs!

¹⁸ *i2 Technologies, Inc. et al v. Oracle Corporation et al.*, Case No. 10-cv-00284 Dkt. Nos. 85 & 130 (E.D.Tex.) (i2 asserted several predictive analytics patents against Oracle); Erin Coe, *I2, Oracle Resolve Software Patent Battle*, LAW360, March 4, 2011, <http://www.law360.com/articles/229787/i2-oracle-resolve-software-patent-battle>.

They're complex programs. *They're not like laws of physics or laws of nature. They're created by us.*¹⁹

23. The '282 patent discloses how interactions with the internet are manipulated to yield a desired result—a result that overrides the routine and conventional sequence of events ordinarily triggered by requesting content or a product that is relevant to a user of a website.

24. And the use of probability values in collaborative filtering (as in the '282 patent) to control for generally popular content and/or products is important and offers something more than a collaborative filtering system that fails to control for the general popularity of content and/or products. Data scientists at Hulu, LLC (operator of a streaming video website) described the importance of accounting for general popularity of a given item:

Just because a recommendation system can accurately predict user behavior does not mean it produces a show that you want to recommend to an active user. For example, “Family Guy” is a very popular show on Hulu, and thus most users have watched at least some episodes from this show. These users do not need us to recommend this show to them — the show is popular enough that users will decide whether or not to watch it by themselves. Thus, *novelty is also an important metric to evaluate recommendations.*²⁰

25. Oracle white papers have identified a critical drawback in many recommendation systems is the failure to account for “popularity bias.”

They tend to over recommend popular items because they are clicked on more often. This would be the equivalent of an in-store sales associate only recommending items on the end-caps and never looking at merchandise on the upper shelves in the aisles—often the items that are most difficult to find.²¹

26. Ten years after Gary Robinson conceived of the inventions in the '282 patent, a 2005 White Paper from Oracle, entitled “The Art of Personalization,” described the use of collaborative filtering to provide recommendations as “new technology” and a “breakthrough.”

¹⁹ *What Makes Algorithms Go Awry?*, NATIONAL PUBLIC RADIO, ALL THINGS CONSIDERED, June 5, 2015, <http://www.npr.org/sections/alltechconsidered/2015/06/07/412481743> (emphasis added) (the quotation is from an interview with Dr. Tufekci).

²⁰ Liang Xiang, Hua Zheng & Hang Li, *Hulu's Recommendation Engine*, HULU TECH BLOG, Sept. 19, 2011, <http://tech.hulu.com/blog/2011/09/19/recommendation-system/> (emphasis added).

²¹ Next-Generation Product Recommendations, Oracle White Papers 6 (March 2011).

Collaborative filtering is relatively *new technology that can deliver better results*. Just go to the leading Web sites that offer “recommendations” and you notice the value. After purchasing a book on *Learning to Golf*, you later return to the Web site and find other books on *Greatest Golf Courses* and *Golf Tips from the Pros*. These recommendations seem relevant, timely, and yet sometimes simplistic. Often you’ll see other *Learn to...* books and videos, like *Learn to Ski*, *Learn to Play Tennis*, and *Learn to Sew*. Compared to past manual attempts at personalization and “e-expectations,” this is a *breakthrough*.²²

THE PARTIES

27. McKinney, Texas based Fellowship Filtering is committed to advancing the current state of technology in the field of predictive analytics systems. In addition to the ongoing efforts of Mr. Robinson, Fellowship Filtering employs a McKinney, Texas resident as a Technology Analyst. Fellowship Filtering is a Texas limited liability company with its principal place of business at 6851 Virginia Parkway, Suite 214, McKinney, Texas.



28. Fellowship Filtering is a small, Texas-based company. Fellowship Filtering depends on patent protection to effectively license its innovative technologies and build its business.

29. On information and belief, Cloudera is a Delaware corporation with its principal office at 220 Portage Avenue, Palo Alto, CA 94306. Cloudera can be served through its registered

²² CHARLES BERGER, ORACLE WHITE PAPER: THE ART OF PERSONALIZATION 4 (August 2005) (emphasis added).

agent, Corporation Service Company d/b/a CSC – Lawyers Incorporating Service Company, 211 E. 7th Street, Suite 620, Austin, TX 78701.

30. On information and belief, Cloudera has offices in Texas where it sells, develops, and/or markets its products including, an office at 515 Congress, Suite 1450, Austin, TX 78701. Further, Cloudera hosts “Cloudera Sessions” at various locations throughout the State of Texas in an effort to market its products.²³

31. According to Cloudera’s website, Cloudera offers infringing products for sale throughout the United States, including in the Eastern District of Texas. Further, Cloudera advertises its infringing products throughout the Eastern District of Texas and claims financial benefits through its conducting of business in Texas. For example, it is registered to do business in the State of Texas, maintains an active office in Austin, Texas, and it actively markets its products by conducting seminars entitled “Cloudera Sessions” at various cities throughout Texas.

32. On information and belief, Cloudera has acquired companies in the field of data analytics that are based in Texas, including Gazzang, Inc. of Austin, Texas.

JURISDICTION AND VENUE

33. This action arises under the patent laws of the United States, Title 35 of the United States Code. Accordingly, this Court has exclusive subject matter jurisdiction over this action under 28 U.S.C. §§ 1331 and 1338(a).

34. Upon information and belief, this Court has personal jurisdiction over Cloudera in this action because Cloudera has committed acts within the Eastern District of Texas giving rise to this action and has established minimum contacts with this forum such that the exercise of jurisdiction over Cloudera would not offend traditional notions of fair play and substantial justice. Defendant Cloudera, directly and through subsidiaries or intermediaries (including distributors, retailers, and others), has committed and continues to commit acts of infringement in this District by, among other things, offering to sell and selling products and/or services that infringe the ‘282

²³ See http://go.cloudera.com/Houston_2014; http://go.cloudera.com/dallas_2014.

patent. Moreover, Cloudera is registered to do business in the state of Texas, and has appointed Corporation Service Company d/b/a CSC – Lawyers Incorporating Service Company, 211 E. 7th Street, Suite 620, Austin, TX 78701, as its agent for service of process.

35. Venue is proper in this district under 28 U.S.C. §§ 1391(b), 1391(c) and 1400(b). Defendant Cloudera is registered to do business in Texas, and upon information and belief, has transacted business in the Eastern District of Texas and has committed acts of direct and indirect infringement in the Eastern District of Texas.²⁴

TECHNOLOGY BACKGROUND

36. Advances in computational power and the explosive growth of the internet have led to the development of data analytics systems for accurately recommending content and products to internet users. The ‘282 patent teaches specific automated collaborative filtering (“Automated CF” or “ACF”) technologies for recommending products and content to users of the internet.

37. Cloudera’s competitors such as SAP have confirmed the importance and value of collaborative filtering systems to providing accurate recommendations.

²⁴ *i2 Technologies, Inc. et al v. Oracle Corporation et al.*, Case No. 10-cv-00284 Dkt. Nos. 85 & 130 (E.D.Tex.).

According to Welington Fonseca, VP of marketing and digital analytics, “Gilt’s commitment to a personalized experience is evident when customers return to the home page of the web site or mobile app. Sales within the store (men, women, kids, home) with the highest affinity to a consumer’s past behavior and preferences (browse, purchase, favorite brands, wish list) are presented at the top of their home page with all other sales ranked according to relevance based on previous shopping behavior and collaborative filtering.”

Another example of personalization is “Your Personal Sale,” which displays the most relevant brands and products based on shopping patterns and self-stated preferences and provide Gilt with another way to interact with the customer to understand preferences in order to further refine their personalization algorithm.

Eman Roman, *Why You Need Human Data for Real Customer Engagement*, SAP BUSINESS INNOVATION BLOG, February 27, 2015, <http://blogs.sap.com/innovation/sales-marketing/why-need-human-data-customer-engagement-02271337> (emphasis added).

38. Although content and product recommendations on websites are commonplace today, at the time the inventions disclosed in the ‘282 patent were conceived, an advanced system for recommending products and content automatically utilizing variables (*e.g.*, multiple users, product ratings, purchase history, and/or actions of website users) was novel.

39. The claims in the ‘282 patent describe a solution that is unquestionably rooted in computer technology to overcome a problem specific to and characteristic of computer networks.

Today increasing numbers of people are turning to computational recommender systems. ***Emerging in response to the technological possibilities and human needs created by the World Wide Web***, these systems aim to mediate, support, or automate the everyday process of sharing recommendations.²⁵

40. The Tapestry system, developed in 1992, introduced the idea (and terminology) of collaborative filtering.²⁶ Tapestry was developed at Xerox’s Palo Alto Research Center and was

²⁵ Loren Terveen & Will Hill, *Beyond Recommender Systems: Helping People Help Each Other*, in *HCI IN THE NEW MILLENNIUM 2* (Jack Carroll, ed., Addison-Wesley, 2001) (emphasis added).

²⁶ David Goldberg, David Nichols, Brian M. Oki, & Douglas Terry, *Using Collaborative Filtering to Weave an Information Tapestry*, COMMUNICATIONS OF THE ACM 35 No. 12, 61–70 (1992) (One of the first uses of the term “collaborative filtering” can be found in this paper.).

directed to electronic mail filtering. Tapestry was based on identifying relevant email content based on exploiting explicit feedback (ratings and annotations) of other users. Tapestry stored the contents of messages, along with metadata about authors, readers, and responders. It allowed any user to store annotations about messages, such as "useful survey" or "Gary should see this!" Tapestry users could form queries that combined basic textual information (*e.g.*, contains the phrase "recommender systems") with semantic metadata queries (*e.g.*, written by Gary OR replied to by Joe) and annotation queries (*e.g.*, marked as "excellent" by Chris).

41. The development of the first collaborative filtering system was directly motivated by the need to sort electronic content transmitted over the internet (*e.g.*, electronic messages posted to newsgroups). "The motivation for Tapestry comes from the increasing use of electronic mail, which is resulting in users being inundated by a huge stream of incoming documents."²⁷

42. Although widely adopted today, in the 1990's, collaborative filtering was a groundbreaking technology offering significant benefits over existing recommendation systems that were content based ("content-based filtering"). Content-based filtering made recommendations based on the content of a document. The creators of Tapestry described this break from prior systems:

Collaborative filtering is *novel because it involves the relationship between two or more documents*, namely a message and its reply, or a document and its annotations. Unlike current filtering systems, Tapestry filters cannot be computed by simply examining a document when it arrives, but rather require (potentially) repeatedly issuing queries over the entire database of previously received documents. This is because sometime after a document arrives, a human (say Smith) may read that document and decide it is interesting. At the time he replies to it (or annotates it), you want your filter to trigger and send you the original document.²⁸

²⁷ *Id.*

²⁸ *Id.* at 61 (emphasis added).

43. Tapestry illustrates the limitations present in systems contemporaneous to the '282 patent. Tapestry lacked the ability to recommend content automatically based on similarities between users. Instead, the Tapestry system worked by recommending content based on predefined filters set by a second user.²⁹ For example, if a user wanted to prioritize messages relating to "Bakersfield, California" the system would return all messages that had previously been "tagged" by prior users as relating to "Bakersfield, California."

44. The below images show the Tapestry system prioritized content based on users requesting content previously tagged by another user of the Tapestry system.

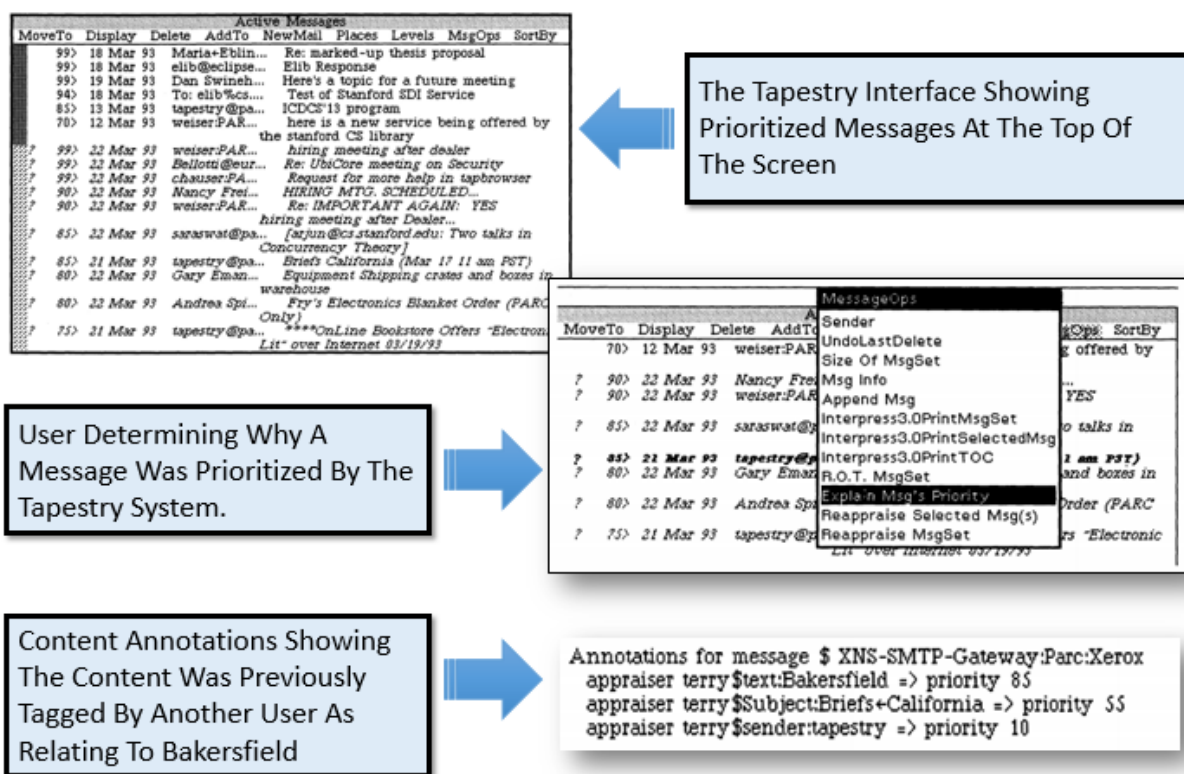


Fig. 1 (images of the Tapestry System (explanation added in blue)).³⁰

45. Another early collaborative filtering system contemporaneous to the '282 patent was GroupLens. Started in 1994 by researchers at the Massachusetts Institute of Technology and

²⁹ The Tapestry system was similar in many ways to Mr. Robinson's earlier 1980's matching system utilized in the Relationship Matching Service.

³⁰ Douglas B. Terry, *A Tour Through Tapestry*, in PROCEEDINGS OF THE CONFERENCE ON ORGANIZATIONAL COMPUTING SYSTEMS 21-30 (Simon Kaplan ed. 2003).

later the University of Minnesota, the GroupLens system implemented a collaborative filtering system for rating Usenet newsgroup articles.³¹ To make personalized predictions identifying the most useful Usenet articles to a user, the GroupLens system asked each user to enter a 1 to 5 rating after reading an article. GroupLens collected the ratings data in a database and compared these ratings to find users who shared similar tastes. Users of GroupLens were then provided a predictive rating for unread Usenet articles. The predictive rating was based on other users who shared similar taste with the user.

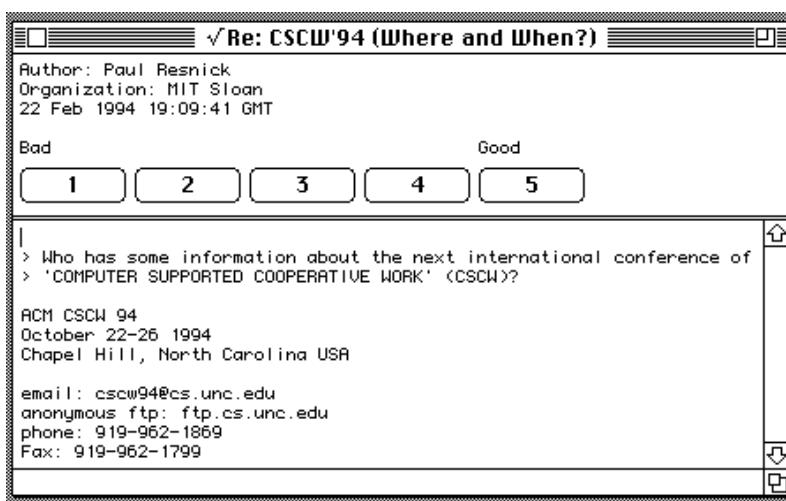


Fig. 2 (showing the user interface for GroupLens and the ability to rate articles 1-5).³²

³¹ Paul Resnick et al., *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*, in PROCEEDINGS OF ACM 1994 CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK 175—86 (1994).

³² *Id.* at Fig. 3.

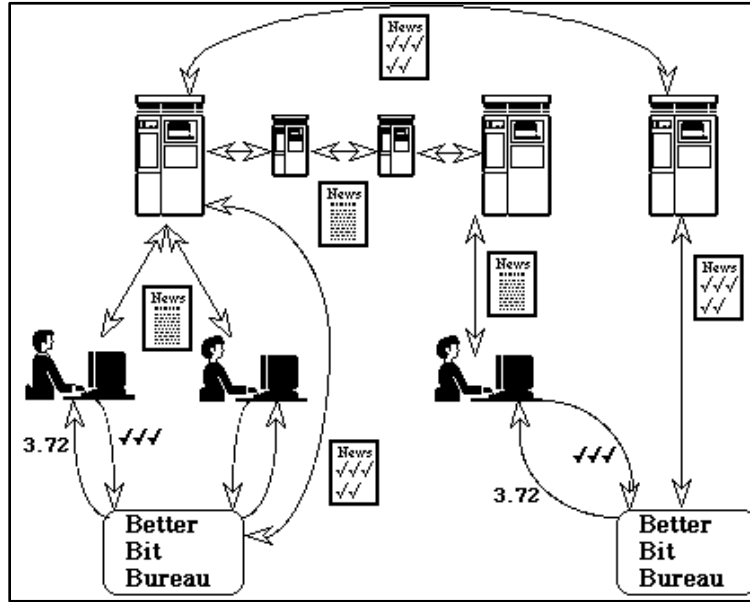


Fig. 3 (showing the architecture of the GroupLens system).³³

46. GroupLens illustrates limitations in automated filtering systems contemporaneous to the '282 patent. The GroupLens system used the Pearson correlation to calculate similarities between users and used the similarities between users to generate predictive ratings. The Pearson correlation coefficient is calculated by comparing ratings for all items rated by both the target user and the neighbor (*e.g.*, correlated items). The equation below gives the formula for the Pearson correlation between user "u" and neighbor "n," where CR_{u,n} denotes the set of correlated items between u and n.

$$userSim(u, n) = \frac{\sum_{i \in CR_{u,n}} (r_{ui} - \bar{r}_u)(r_{ni} - \bar{r}_n)}{\sqrt{\sum_{i \in CR_{u,n}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in CR_{u,n}} (r_{ni} - \bar{r}_n)^2}}$$

47. The Pearson correlation and contemporaneous systems to the '282 patent failed to incorporate agreement about content in the population as a whole. For instance, the system failed to account for the fact that two users' agreement about a universally loved movie was less important than agreement on a controversial or unpopular movie. The Pearson correlation failed to

³³ *Id.* at Fig. 2.

capture distinctions relating to an item's general popularity. Thus, GroupLens made predictions based on data that showed similarities (arising from a piece of content being generally popular) but GroupLens' recommendations were not statistically significant.

48. John Hey's patents (U.S. Pat. Nos. 4,996,642 and 4,870,579), which are cited on the face of the '282 patent, describe a system for recommending items based on ratings of the items. Like GroupLens and other systems contemporaneous to the '282 patent, Hey's system for recommending products based on user ratings failed to account for statistically significant similarities between certain users; the recommendations were merely the product of an item or piece of content being generally popular. This prevented the Hey system from offering accurate predictions and recommendations of items and content.

49. Similarly, the Ringo music recommendation system, discussed by Upendra Shardanand and Pattie Maes, and cited on the face of the '282 patent, used Pearson's correlation measure to provide content and product recommendations. Like other systems contemporaneous to the '282 patent, Shardanand and Maes's system failed to take into account the statistically significant similarities between certain users.³⁴ Information showing unusual similarity in preferences for particular users was unutilized. Furthermore, these prior art systems did not provide recommendations with statistically meaningful confidence levels as the number of items that both the user and a respective recommending user provided ratings for increased.

50. Collaborative filtering arose to solve problems faced by digital content providers in the internet era, as described by Adobe's Global Alliance Manager, Jamie Brighton:

The catalyst for the evolution of personalization has been competition through, a product of the Internet's explosive growth. This growth provided consumers with so many options for e-commerce that it created a market in desperate need of a process by which consumers could develop a personal

³⁴ Upendra Shardanand & Pattie Maes, *Social Information Filtering: Algorithms for Automating Word of Mouth*, in PROCEEDINGS OF CHI '95 CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 210—17 (1995).

connection with a brand or digital storefront in a sea of rapidly evolving competitors.³⁵

51. At the time the inventions disclosed in the '282 patent were conceived, the internet and the state of technology generally was vastly different from 2015, or even the state of the internet 10 years ago. For example, Facebook.com, YouTube.com, Wikipedia.com, and LinkedIn.com were years from being launched.³⁶

³⁵ Jamie Brighton, *Changes in Personalization and What's Coming Next*, ADOBE DIGITAL MARKETING BLOG, October 21, 2014, <http://blogs.adobe.com/digitalmarketing/personalization/personalization-past-present-future/>.

³⁶ Rob Waugh, *Before They Ruled the Internet: 'Ancient' Home Pages for Amazon, Google and 'The Facebook' Show Much Web Giants Have Changed*, DAILY MAIL, January 19, 2012, <http://www.dailymail.co.uk/sciencetech/article-2088445>; TONY SEBA, WINNERS TAKE ALL – THE 9 FUNDAMENTAL RULES OF HIGH TECH STRATEGY 137 (2006); GEORGE A BARNETT, ENCYCLOPEDIA OF SOCIAL NETWORKS 947 (2011).



The above images show major internet properties contemporaneous (and later) to the inventions conceived in the '282 patent, including: Google.com (September 1998), Yahoo.com (March 1995), Amazon.com (1995), Myspace.com (August 2003).³⁷

52. Academics such as Daniela M. Witten of the University of Washington describe the development of collaborative filtering systems as directed to solving problems arising out of so called Big Data (a term for modern networked computers that capture considerable volumes of data).

Collaborative filtering is one example of a statistical method that has been newly-developed in the context of Big Data, in order to answer a question that didn't arise with Small Data. Collaborative filtering systems are used by companies like Amazon to suggest to a customer items that he or she might

³⁷ *Id.*

want to purchase, based on his or her past purchase history as well as purchases made by other customers.³⁸

53. Collaborative filtering systems, such as the system taught in the '282 patent were directed to solving a problem unique to the internet using uniquely computer based technologies.

Computers and the web allow us to advance beyond simple word-of-mouth. Instead of limiting ourselves to tens or hundreds of individuals the Internet allows us to consider the opinions of thousands. The speed of computers allows us to process these opinions in real time and determine not only what a much larger community thinks of an item, but also develop a truly personalized view of that item using the opinions most appropriate for a given user or group of users.

J. Ben Schafer, Dan Frankowski, Jon Herlocker & Shilad Sen, *Collaborative Filtering Recommender Systems*, in THE ADAPTIVE WEB: METHODS AND STRATEGIES OF WEB PERSONALIZATION 292 (Peter Brusilovsky *et al.* eds., 2007).

54. On information and belief, contemporaneous to, and following Mr. Robinson's conception of the inventions disclosed in the '282 patent, academics, and businesses headquartered in Texas actively entered the field of collaborative filtering. Computer researchers at the University of Texas at Austin founded the Intelligent Data Exploration and Analysis Laboratory and the Machine Learning Research Group. The University of Texas at Dallas founded the Institute of Data Analytics, a center for research on data analysis, which collaborates with private industry. Baylor University in Waco, Texas is the home of the Electronic Commerce Center, which focuses on integrating technology and electronic data with e-commerce.

55. Texas based companies incorporated collaborative filtering technologies into numerous products and many of these same companies cited the '282 patent in their own patents. Texas based businesses that developed products incorporating collaborative filtering included: VideosDotCom, Inc. of McKinney, Texas; i2 Technologies US, Inc. of Dallas, Texas; Vignette Corporation of Austin, Texas; Texas Shopper Network, Inc. of Houston, Texas; Arrowsmith Technologies, Inc. of Austin, Texas; and HP Enterprise Services, LLC of Plano, Texas.

³⁸ Nicholas Bashour, *The Big Data Blog, Part II: Daniela Witten*, AAAS NEWS, March 17, 2014, <http://www.aaas.org/news/big-data-blog-part-ii-daniela-witten>.

56. The '282 patent is cited by at least 60 patents that were either initially assigned to or are currently assigned to entities headquartered in Texas. Companies citing the '282 patent in their patents include i2 Technologies, Vignette Corporation, AT&T, Hewlett-Packard Development Company, and Blockbuster LLC.

THE VALUE OF MR. ROBINSON'S INVENTION

57. Executives at leading technology companies have described the value of accurate product and content recommendations as critical, lasting, and prominent. Jamie Brighton, Global Alliance Manager at Adobe, stated accurate recommendation techniques were “a light switch for innovators and marketers alike, as well as a warning. A warning that personalization was rapidly becoming the ultimate avenue for creating lasting partnerships with a digital consumer base, and that ignoring this technology simply wouldn't be an option forever.”³⁹

58. An IBM developerWorks® paper described the importance of providing accurate recommendations.

Recommendation systems changed the way inanimate websites communicate with their users. Rather than providing a static experience in which users search for and potentially buy products, recommender systems increase interaction to provide a richer experience. Recommender systems identify recommendations autonomously for individual users based on past purchases and searches, and on other users' behavior.⁴⁰

59. Numerous companies have confirmed the value of providing accurate product recommendations. “By showing the visitor the content they are looking for, you increase conversion rates and reduce bounce rates.”⁴¹ Companies such as HP, RichRelevance, and Adobe

³⁹ Jamie Brighton, *Changes in Personalization and What's Coming Next*, ADOBE DIGITAL MARKETING BLOG, October 21, 2014, <http://blogs.adobe.com/digitalmarketing/personalization/personalization-past-present-future/>.

⁴⁰ M. Tim Jones, *IBM Developer Works: Recommender Systems, Part 1: Introduction to Approaches and Algorithms 2* (December 12, 2013), available at <http://www.ibm.com/developerworks/library/os-recommender1/>.

⁴¹ *Cognitor: Content Guidance and Recommendations 2*, COGNITOR WEBSITE, April 15, 2015, <http://www.cognitor.com/brochures/enterprise.pdf>.

confirm the importance of collaborative filtering technologies to generating accurate recommendations.

With these concerns in mind, RichRelevance based the enRICH platform on multiple recommendation strategies, ranging from simple categorical top sellers, to collaborative filtering algorithms After deploying the enRICH platform, retail customers report improvements across a range of KPIs, including increased conversion, revenue, and repeat visits.⁴²

In its simplest form, collaborative filtering really works when data from multiple sources comes together and is sorted into categories. ***It is a must these days*** for any e-commerce site striving to deliver a basic level of website personalization.⁴³

Personalized services are becoming increasingly indispensable on the Web, ranging from providing search results to product recommendation. Examples of such systems include recommending products at Amazon.com, DVDs at Netflix, News by Google etc. The central technique used in these systems is collaborative filtering (CF) which aims at predicting the preference of items for a particular user based on the items previously rated by all users.⁴⁴

The truth is indisputable—optimization increases conversion, so every digital property needs optimization. This singular truth is transforming the practice of marketing. Now, marketers must tap into the constant stream of web activity and customer data to gain insight into what visitors and customers want to see and experience. ***They must immediately act on that insight and deliver highly relevant, personalized content*** throughout the customer life cycle.⁴⁵

⁴² Rich Relevance, *Speak <geek> [sic] Technical Brief 6* (2009), available at http://www.richrelevance.com/wp-content/uploads/2011/01/Speak-Geek2_EnsembleLearning_RichRelevance.pdf.

⁴³ Dan Darnell, *Collaborative Filtering and Its Importance to Personalized Recommendations in eCommerce*, BAYNOTE BLOG, April 18, 2013, <http://www.baynote.com/2013/04/how-collaborative-filtering-impacts-product-recommendations/> (emphasis added).

⁴⁴ Rong Pang et al., *One-Class Collaborative Filtering*, in IEEE INTERNATIONAL CONFERENCE ON DATA MINING (ICDM 2008) 502—11 (2008) (Mr. Pang at the time was employed by Hewlett-Packard.).

⁴⁵ Adobe Target Premium Overview 1 (2014), available at <http://www.adobe.com/content/dam/Adobe/en/solutions/testing-targeting/pdfs/target-premium-overview-ue.pdf> (emphasis added).

Dynamic, relevant content is proven to increase engagement and conversions by as much as 6 times when compared to static content.⁴⁶

U.S. PATENT NO. 5,885,282

60. Fellowship Filtering is the owner by assignment of the ‘282 patent. The ‘282 patent is entitled "Automated Collaborative Filtering System." The ‘282 patent issued on March 16, 1999, based on a patent application filed on April 9, 1998, and claims priority to a provisional application filed on April 30, 1996. A true and correct copy of the ‘282 patent is attached hereto as Exhibit A.

61. The claims in the ‘282 patent are directed at a unique computing solution that addresses a problem particular to computer networks – the recommendation of items or content based on prior user actions.

62. Recommending content over a computer network presented new and extraordinary issues over the techniques and systems known in the art at the time. Prior art recommendation systems had a number of drawbacks. Such systems “fail to take into account the probability that a random user will provide a given rating. Thus, information showing unusual similarity in preferences for particular users is not utilized.” ‘282 patent, cols. 1:67-2:4.

63. The recommendation technologies claimed in the ‘282 patent were aimed at solving problems specific to the internet. “The catalyst for the evolution of personalization has been competition though, a product of the Internet’s explosive growth. This growth provided consumers with so many options for e-commerce that it created a market in desperate need of a process by which consumers could develop a personal connection with a brand or digital storefront in a sea of rapidly evolving competitors.”⁴⁷

⁴⁶ *BaynoteOne Product Recommendations 1* (204), available at <http://www.baynote.com/wp-content/uploads/2012/04/BaynoteONE-Solution-Brief-Personalized-Product-Recommendations.pdf>.

⁴⁷ Jamie Brighton, *Changes in Personalization and What’s Coming Next*, ADOBE DIGITAL MARKETING BLOG, October 21, 2014, <http://blogs.adobe.com/digitalmarketing/personalization/personalization-past-present-future/>.

64. The technology of “[c]ollaborative filtering is a relatively young algorithmic approach” and thus was not a conventional business practice.⁴⁸

65. One or more claims in the ‘282 patent recite a “similarity calculation.” This element of the ‘282 patent is one of the “inventive concepts” of the ‘282 patent. The use of a similarity calculation is an “inventive concept” allowing computer servers configured to operate websites to more efficiently and accurately recommend content and products to website users.

66. The ‘282 patent does not preempt every way of “providing recommendations using a computer system,” as systems for doing so existed before this invention, and systems exist now that allow website operators to provide recommendations without infringing the claims of the ‘282 patent.

67. The ‘282 patent claims do not preempt the field or preclude the use of other effective recommendation technologies. The ‘282 patent claims include inventive elements such as the use of probability calculations, randomized transformed ratings data, and/or similarity values to generate preference data over a computer network. The elements in the ‘282 claims greatly limit the breadth of the ‘282 patent's claims. These limitations are not necessary or obvious tools for achieving the generation of user preference data and/or recommendations, and they ensure that the claims do not preempt the field of recommendation systems and/or collaborative filtering.

68. Other techniques for collaborative filtering that are not included within the scope of the ‘282 patent's claims include, but are not limited to, the prior art discussed in the ‘282 patent:

- U.S. Patent No. 4,870,579 to Hey teaches providing recommendations to a user based on a user selected from a group of users, the reactions of the selected user to items sampled by one or more users in the group but not sampled by the selected user.
- U.S. Patent No. 4,996,642 to Hey teaches providing recommendations to a user based on other items previously sampled by that user and on

⁴⁸ Yehuda Koren, *Tutorial on Recent Progress in Collaborative Filtering*, in PROCEEDINGS OF THE 2008 ACM CONFERENCE ON RECOMMENDER SYSTEMS (RECSYS '08) 333-334 (2008).

the availability of the item. Further, the recommendations were represented by a scalar rating for each item.

- U.S. Patent No. 5,452,410 to Magidson teaches apparatus and methods for achieving statistical analysis of categorical and continuous outcomes and for displaying the results of such analyses.
- Upendra Shardanand, "Social Information Filtering for Music Recommendation" Sep. 1994, pp. 1-93, Massachusetts Institute of Technology, Thesis. This system attempted to provide recommendations to a user based on ratings for items provided by the user as compared with other users.

69. The '282 patent claims do not preempt the field of recommendation systems. Technologies falling outside the scope of the '282 patent may include, but are not limited to, the following: (1) filtering relying solely on content-based techniques, (2) collaborative filtering using only a standard *Pearson r* correlation coefficient, (3) collaborative filtering relying on the Mean Squared Difference, and (4) community-based recommendation systems.

70. In contrast to the '282 patent, the patents at issue in *I/P Engine Inc. v. AOL Inc.*, claimed all instances of recommendation systems where content and collaborative filtering was used. Judge Mayer, in his Federal Circuit concurring opinion wrote, "the scope of the claimed invention is staggering, potentially covering a significant portion of all online advertising." *I/P Engine, Inc. v. AOL Inc.*, 576 F. App'x 982, 995 (Fed. Cir. 2014). Further, despite the asserted patents (U.S. Patent Nos. 6,314,420 and 6,775,664 ("I/P Engine Patents")) claiming a priority date of 1998 (*Id.* at 997) and a specification 50% shorter than that of the '282 patent, the I/P Engine Patents' broad claims were upheld by the Patent and Trademark Office in two reexamination proceedings, by a jury following a 12 day trial, and by United States District Judge Raymond Alvin Jackson following significant post-trial briefing. In contrast, the provisional application to which the '282 patent claims priority precedes the I/P Engine Patents' priority date by two years and contains significantly narrower claims.

71. The '282 claims are not directed to any "method of organizing human activity," "fundamental economic practice long prevalent in our system of commerce," nor "a building block of the modern economy." Instead, the '282 patent's claims are limited to the realm of systems

utilized in “calculating similarity values” and “recommending products and content” over a “computer network.”

72. The ‘282 patent’s claims are not directed at the broad concept or idea of “recommending items.” Instead, the claims are directed to particular, narrow methods and systems for “providing recommendations by transforming user data,” using technologies unique to the internet age. The inventive concept in the ‘282 claims is a technological one rather than an entrepreneurial one – the development of systems and methods used to calculate content and/or product recommendations that are statistically significant, thus improving the accuracy of the content and/or product recommendations.

73. The ‘282 patent does not take a well-known or established business method or process and “apply it to a general purpose computer.” Instead, the specific system and processes described in the ‘282 patent have no direct corollary to a business process that predates the advent of the internet.

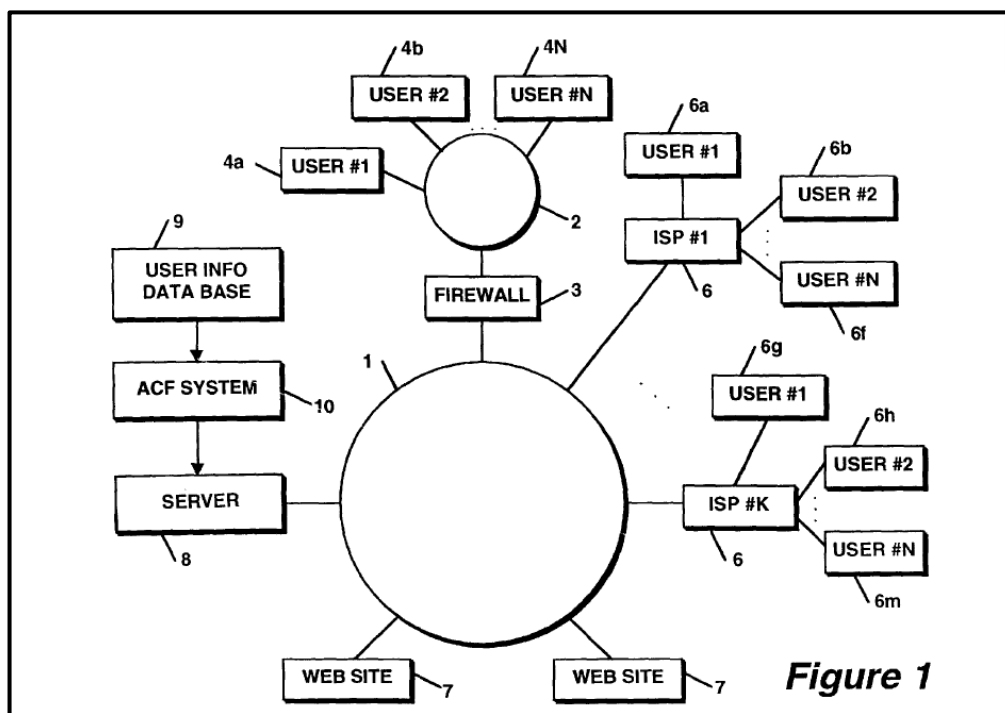
74. The ‘282 patent’s claims are directed toward a solution rooted in computer technology and uses technology unique to computers and networks to overcome a problem specifically arising in the realm of making product and content recommendations over a computer network. For example, the ‘282 patent’s claims are directed toward generating recommendations using data collected in a database from users over the internet — a result that overrides the routine and conventional sequence for providing recommendations known in the art at the time the inventions disclosed in the ‘282 patent were conceived.

75. The ‘282 patent’s claims are not directed at a mere mathematical relationship or formula as the ‘282 patent's claims teach specific systems and methods for providing recommendations of content and products over a computer network using both data from prior users of a website as well as information created by the systems and methods described in the ‘282 patent’s claims.

76. The ‘282 patent’s claims cannot be performed by a human, in mind, or by pen and paper. The claims as a whole are directed to generating user preference data using a connection to

the internet to gather data from users, a database to store user data, and a computer processor to conduct complex statistical calculations. These limitations establish that the '282 patent's claims are not an abstract idea, because they cannot be performed by a human, in the human mind, or by pen and paper.

77. Further, the '282 patent disclosure requires a computer to generate content and/or product recommendations. For example, in block 90, the method disclosed in the '282 patent computes whether the similarity value is sufficient to generate preference data. The result of the steps described in the '282 patent is a computer server using processing power to conduct complex calculations over large data sets and creating new data used by the system to improve the quality of



recommendations.

Fig. 4 (showing the implementation of the '282 patent system arose from receiving user data over the internet including through a website).⁴⁹

78. The use of probability calculations to generate user preference data is not a

⁴⁹ '282 patent, fig. 1.

conventional, routine activity in which humans engage.

79. The prior art cited on the face of the '282 patent further shows the invention claimed in the '282 patent is not a patent ineligible abstract idea. The invention described in the '282 patent's claims is narrower than much of the cited prior art, and therefore, is not an abstract idea. For example, U.S. Pat. Nos. 4,996,642 to Hey describes systems and methods that attempted to provide recommendations to a user based on ratings for items provided by the user as compared with other users. The '282 patent's claims require additional limitations and thus the '282 patent's claims are directed toward significantly more than an abstract idea and the '282 patent's claims do not preempt the field of recommendation engines or even collaborative filtering.

80. The claimed invention in the '282 patent's claims is rooted in computer technology and overcame a problem specifically arising in the realm of computer networks. The '282 patent's claims require the use of a computer system.

81. The use of a computer system plays a significant part in performing the claims of the '282 patent. For example, the use of a computer processor to generate user preference data utilizing data stored in a computer database is integral to the success of the system, and can only be performed using a computer system. The use of a computer system to process user data stored in a database does far more than improve the efficiency of the process; the computer system is integral to accomplishing the generating of recommendation data.

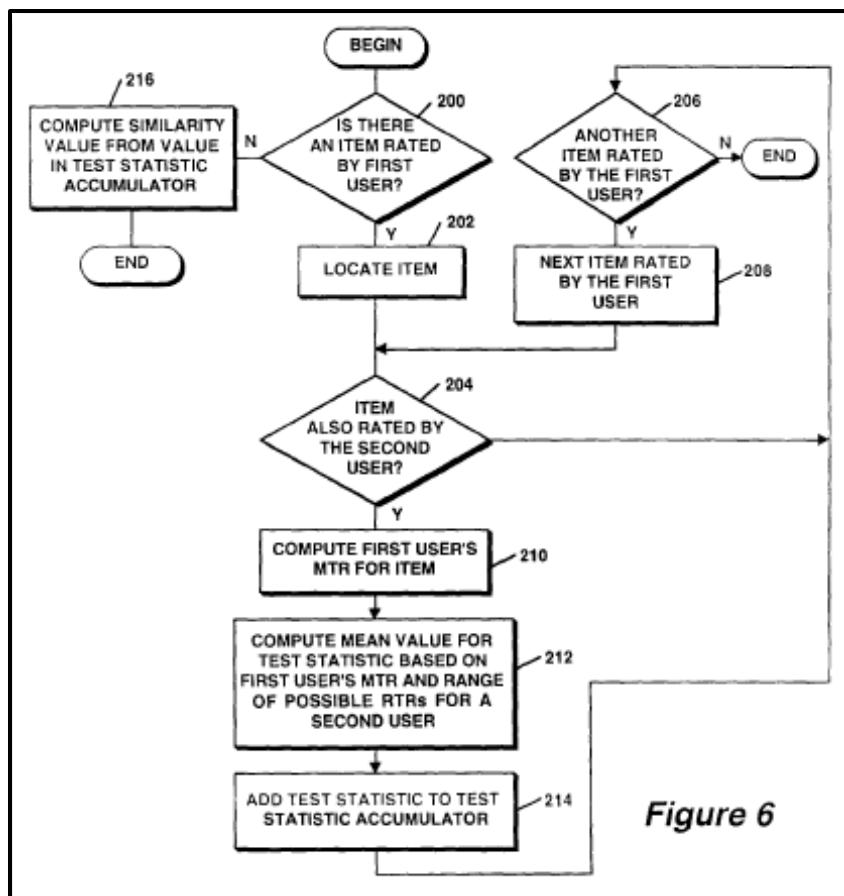


Fig. 5 (showing the generation of recommendation data).⁵⁰

82. The rising volume of content and data made possible by the internet drives the need to identify relevant products and content using filtering technologies such as that disclosed in the '282 patent.

With the development and popularity of WWW, billions of web pages are retrievable via search engines like Google. Despite it is not a perfect method to find what we want, most search engines still use keywords in documents and queries to calculate the relevance. As the only interface for users accessing tremendous web pages, queries are one of the most important factors that affects the performance of search engines. However, web pages returned from search engines are not always relevant to user search intentions. An independent survey of 40,000 web users found that after a failed search,

⁵⁰ '282 patent, fig. 6.

76% of them will try to rephrase their queries on the same search engine instead of resorting to a different one.⁵¹

83. Dan Darnell, a Senior Director of Product Marketing at Baynote, similarly described collaborative filtering as directed to solving problems specific to the internet:

In its simplest form, collaborative filtering really works when data from multiple sources comes together and is sorted into categories. It is a must these days for any e-commerce site striving to deliver a basic level of website personalization.⁵²

84. Academics have recognized that the development of collaborative filtering recommendation systems is directly tied to and an outgrowth of information overload problems created by and unique to the internet.

The challenge of finding the needed information from the web has led to the development of a number of recommender systems, which typically watch the user navigation behavior as a sequence of pages and suggest another set of web pages, products and other information besides the actual information. With the exponential growth of the web, the study of modeling and predicting a user's access on the web has become crucial to the researchers and portal developers.⁵³

To overcome this so called "information overload" problem, in the mid-1990s researchers started to investigate recommender systems. A recommender system (RS) uses knowledge about your preferences (and those of others) to recommend items you are likely to enjoy. Users can offer feedback on items they are familiar with for example, and the recommender system uses the information to predict their preference for yet unseen items and subsequently recommends items with the highest predicted relevance.⁵⁴

⁵¹ Zhiyuan Liu & Maosong Sun, *Asymmetrical Query Recommendation Method Based on Bipartite Network Resource Allocation*, in PROC. OF WWW'08 1049 (2008).

⁵² Dan Darnell, *Collaborative Filtering and Its Importance to Personalized Recommendations in eCommerce*, INTELLIGENCE COLLECTED: THE BAYNOTE BLOG, April 18, 2013, <http://www.baynote.com/2013/04/how-collaborative-filtering-impacts-product-recommendations/> (Dan Darnell is a Senior Director of product marketing at Baynote).

⁵³ Gopinath Ganapathy & P.K. Arunesh, *Feature Analysis of Recommender Techniques Employed in the Recommendation Engines*, J. COMPUT. SCI. 6(7): 748—55 (2010).

⁵⁴ Joost de Wit, *Evaluating Recommender Systems -- An Evaluation Framework to Predict User Satisfaction for Recommender Systems in an Electronic Program Guide Context* 9 (May 2008), Master's thesis, University of Twente, <http://essay.utwente.nl/59711/>.

85. A 2009 paper supported by the Samsung Research Fund, ties collaborative filtering technologies to solving problems unique to the internet – the generation of information using a common communications infrastructure.

The amount of information on the Web is increasing according to the growth of information and communication infrastructure. As a result, recommender systems (RSs) for personalization are required. An RS provides content or items considering the tastes of individual users. Among the various RSs, collaborative filtering (CF) is the process of filtering for information or patterns using collaborative techniques involving multiple users.⁵⁵

86. Years after the Ringo system was developed (the Ringo system is referenced on the face of the ‘282 patent), the use of collaborative filtering techniques was described as “innovative” by data scientists.

Ringo also provides an innovative solution that inverts the basic CF approach; music albums are treated as ‘participants’ that can recommend users to other music album participants.⁵⁶

87. One or more of the ‘282 patent’s claims relate to a computer-implemented method to transform website user data in a particular manner – by inserting information into user data and using the code to recommend content and/or products. This insertion enables the computer system to recommend content and/or products and generate similarity values.

88. One or more of the claims in ‘282 patent go beyond manipulating, reorganizing, or collecting data by actually adding information associated with a user and using that information to generate a recommendation of a product or content over a computer network, thereby fundamentally altering ratings data associated with a user.

89. One or more of the claims in the ‘282 patent require ‘transforming’ data to generate “randomized ratings data” by “adding a uniformly distributed random number to the ratings data provided by the plurality of users.” Therefore, the claims in the ‘282 patent alter data associated

⁵⁵ Hyeon-Joon Kwon et al., *Improved Memory-based Collaborative Filtering Using Entropy-based Similarity Measures*, in SYMPOSIA AND WORKSHOPS ON UBIQUITOUS, AUTOMATIC AND TRUSTED COMPUTING (WISA’09) (May 2009) (this work was supported by Samsung).

⁵⁶ Sonny Han Seng Chee et al., *Rectree: An Efficient Collaborative Filtering Method*, in 3RD INT. CONF. ON DATA WAREHOUSING AND KNOWLEDGE DISCOVERY (DAWAK 2001) 141 (2001).

with a user and go beyond the mere collection, organization, manipulation, or reorganization of data. The claimed invention goes beyond manipulating, reorganizing, or collecting data by actually adding a new subset of numbers or characters to the data, thereby fundamentally altering the original information.

90. One or more of the claims in the ‘282 patent requires ‘transforming’ one thing (‘ratings data’) ‘to create’ something else (‘randomized ratings data’) and further recites a particular manner of transforming (‘by adding a uniformly distributed random number to the ratings data provided by the plurality of users’). Therefore, claimed features in the ‘282 patent “fundamentally alter” data or “transform” the data.

91. Nor does collaborative filtering merely “support an existing activity.” Professor Loren G. Terveen of the University of Minnesota⁵⁷ and Will Hill of AT&T Labs described collaborative filtering as improving the functioning of computer-based recommendation systems by updating a computer database and transforming data.

Collaborative filtering does not simply support an existing activity. Instead, it requires users to engage in a somewhat novel computationally mediated activity. This activity has a single combined role, the recommendation seeker / preference provider. We describe this as *role uniformity*. Everyone does the same work (rates items) and receives the same benefits (gets rated items as recommendations). We might describe rating items as an “ante” – to get recommendations, you have to give them. ***This leads naturally to growth in the system’s knowledge (and thus to better recommendations), since using the database leads to the database being updated.***⁵⁸

92. White papers from various corporations describe computer-implemented recommendation systems as transforming the data of a previously static website – generating preference information that previously did not exist. Recommendation systems like the inventions disclosed in the ‘282 patent utilize a system for modifying data that has a concrete effect in the field of website and internet usage.

⁵⁷ Loren Terveen was a principal member of the technical staff at AT&T Labs.

⁵⁸ Loren Terveen & Will Hill, *Beyond Recommender Systems: Helping People Help Each Other*, in HCI IN THE NEW MILLENNIUM 13 (Jack Carroll, ed., Addison-Wesley, 2001) (emphasis added).

Rather than providing a static experience in which users search for and potentially buy products, recommender systems increase interaction to provide a richer experience. Recommender systems identify recommendations autonomously for individual users based on past purchases and searches, and on other users' behavior.⁵⁹

93. Further, the '282 patent claims improve upon the functioning of a computer system. "Performance improves as the number of entries in the database increases." '282 patent, col. 23:29-30. The claims and specification of the '282 patent also describe the use of "cluster analysis," which improves the functioning of a computer handling the making of recommendations. "As a means for more efficient processing, cluster analysis can be used." *Id.* 20:36-37.

94. One or more of the claims of the '282 patent recite a means or step for performing a specified function. The corresponding structure(s) in the '282 patent specification and appendix include computer code that improves the functioning of a computer by being more "RAM-efficient." '282 patent, cols. 33:1-39:60.

95. Academic research has confirmed that using ratings improves the functioning of a computer conducting collaborative filtering.

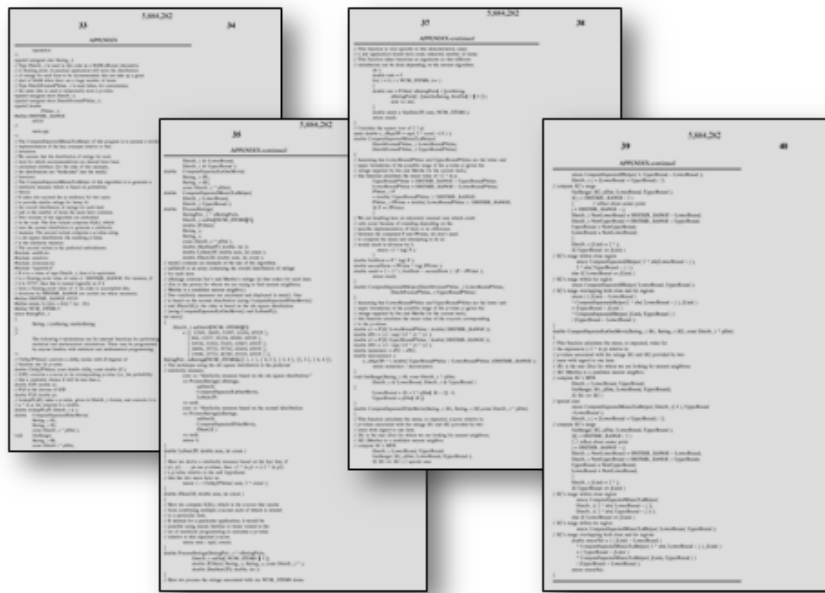
One way to make recommendations of regular, but interesting items, more likely consists in assigning weights to items that devalue ratings given to popular items and appreciate ratings given to regular items. . . . The results of the first set of experiments are shown in Fig. 5. The precision@n values show that when using the weighting functions, the resulting precision@n is slightly higher for low values of n than for the unweighted approach for the Moviepilot dataset (n=5). For the Movielens dataset, the unweighted approach seems to have the upper hand. However, as n increases, the improvement decreases and at a relatively large n (n=50) the weighted approaches perform worse than the non weighted one. In the Movielens case, the unweighted approach always outperforms the weighted ones, irrelevant of n's value. This seems to be in agreement with the findings by Herlocker et al.

⁵⁹ M. Tim Jones, *IBM Developer Works: Recommender Systems, Part 1: Introduction to Approaches and Algorithms 2* (December 12, 2013), available at <http://www.ibm.com/developerworks/library/os-recommender1/>.

Results for the Euclidean and cosine measures showed very similar trends and have thus been omitted.⁶⁰

96. One or more of the claims in the '282 patent recite means-plus-function claim limitations governed by 35 U.S.C. § 112, ¶ 6.

97. The '282 patent discloses computer algorithms in an appendix to the specification. In addition to the structures and algorithms disclosed throughout the specification, these algorithms correspond to means-plus-function claims in the '282 patent.



'282 patent, cols. 39-40 (computer algorithms disclosed in an appendix to the specification).

98. Means-plus-function claims such as those included in the '282 patent are inherently not abstract ideas. Stanford Law Professor Mark Lemley described his analysis:

If the patent is interpreted as a means-plus-function claim, it will be limited to the particular software implementation the patentee actually built or

⁶⁰ Alan Said et al., *Analyzing Weighting Schemes in Collaborative Filtering: Cold Start, Post Cold Start and Power Users*, in PROCEEDINGS OF THE 27TH ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING (SAC'12) 2035, 2039 (2012).

described. Such a narrow, specific claim should not be an unpatentable “abstract idea.”⁶¹

But if you wrote it [an algorithm] and you included it in the step I think you could survive the *Aristocrat* line of cases and then the question will become well what does equivalent thereof mean? Can I show you my algorithm and say, yeah, this is the approach I took but these other four approaches are equivalent and a computer programmer would look at those and say I don’t care which one of those you use. *And if you can do that then you might end up with a claim that’s still pretty broad even though it’s in means plus function format.*⁶²

COUNT I
INFRINGEMENT OF U.S. PATENT NO. 5,885,282

99. Fellowship Filtering references and incorporates by reference the preceding paragraphs of this Complaint.

100. Cloudera makes, uses, sells, and/or offers for sale in the United States products and/or services for generating product and/or content recommendations.

101. On information and belief, Cloudera recommendation products and/or services provide or support generating product and/or content recommendations based on enhanced collaborative filtering technologies to drive more successful and relevant recommendations.

102. Cloudera sells, markets, and/or distributes Cloudera Enterprise (“Cloudera Enterprise”).

103. On information and belief, Cloudera operates the internet site:
<http://www.cloudera.com/content/cloudera/en/resources.html> (“Cloudera Help Center”).

104. On information and belief, the Cloudera Help Center provides product guides and service manual relating to Cloudera Enterprise.

105. On information and belief, the Cloudera Help Center provides documentation for Cloudera Enterprise.

⁶¹ Mark A. Lemley, *Software Patents and the Return of Functional Claiming*, 2013 WISC. L. REV. 905 (2013).

⁶² Eugene Quinn, *The Ramifications of Alice: A Conversation with Mark Lemley*, IPWATCHDOG BLOG, September 4, 2014, <http://www.ipwatchdog.com/2014/09/04/the-ramifications-of-alice-a-conversation-with-mark-lemley/id=51023/> (emphasis added).

106. On information and belief, Cloudera builds and offers to customers recommendation products and services, such as, Cloudera Enterprise and all versions and variations thereof since the issuance of the '282 patent ("Cloudera Products").

107. On information and belief, Cloudera Products incorporate collaborative filtering technology.

108. On information and belief, Cloudera Products enable the calculation of recommendations based on similarity, so people who bought this bought that or people who viewed this bought that, or people who viewed this viewed that are recommended relevant content or products.

109. On information and belief, Cloudera Products are available to businesses and individuals throughout the United States.

110. On information and belief, Cloudera Products are provided to businesses and individuals located in the Eastern District of Texas.

111. On information and belief, Cloudera conducts trainings relating to Cloudera Products in the Eastern District of Texas.

112. On information and belief, Cloudera Products calculate recommendations based on "explicit ratings" of content and/or products.

113. On information and belief, Cloudera has described its recommendation products as automatically discovering individualized behavior patterns to generate highly accurate recommendations in real time.⁶³

114. On information and belief, Cloudera Products incorporate "RecommenderJob" functionality.

⁶³Josh Patterson, *Recommendation with Apache Mahout in CDH3*, CLOUDERA BLOG, November 30, 2011, <http://blog.cloudera.com/blog/2011/11/recommendation-with-apache-mahout-in-cdh3/> ("With user-based recommendation algorithms we're looking at the notion of similarity between users based on preferences/actions/ratings of those users so we can recommend the same things to similar user.").

115. On information and belief, Cloudera Products enable the calculation of a “weighted sum.”

116. On information and belief, Cloudera Products enables the visualization of correlations between data.

117. On information and belief, Cloudera Products incorporate functionality for calculating Chi-square tests, measures of association, and risk differences.

118. On information and belief, Cloudera Products enable the analysis of distribution values from data sets.

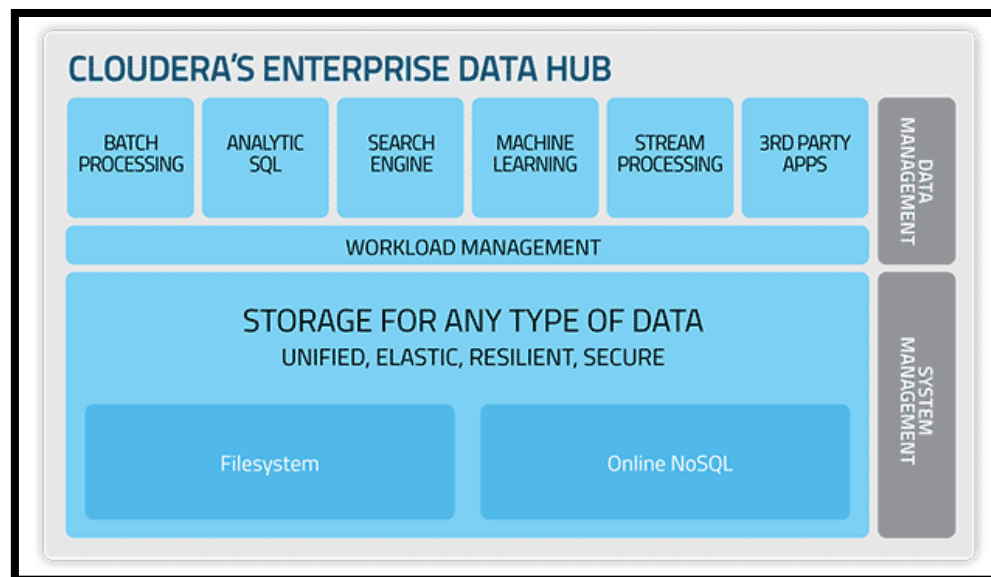
119. On information and belief, Cloudera Products include Mahoot functions.

120. On Information and belief, Cloudera Products perform correlation matrix computation.

121. On information and belief, the recommendation engine in the Cloudera Products support retail business processes by providing recommendations.

122. On information and belief, Cloudera Products enable “Market Basket Analysis.”

123. On information and belief, the following figure shows the architecture of a system implementing Cloudera’s system.



Cloudera Enterprise Background, CLOUDERA WEBSITE, <http://www.cloudera.com/content/cloudera/en/products-and-services/cloudera-enterprise.html> (last visited June 16, 2015).

124. On information and belief, the Cloudera Product enables the “InverseUserFrequency” function.

125. On information and belief, the “InverseUserFrequency” function enables the calculation of a ratings data for content and/or products across all ratings that have been provided to a system.

126. On information and belief, the Cloudera Products enable the computation of a scaling factor “for each item by dividing the total number of users by the number of users expressing a preference for that item, and taking the log of that value. The log base of this calculation can be controlled in the constructor.”⁶⁴

127. On information and belief, the “InverseUserFrequency” function generates a transformed data that provides information on the data distribution within a dataset.

128. On information and belief, one or more of the Cloudera Products generates recommendations by controlling for “popularity bias.”

129. On information and belief, Cloudera Products generate recommendations using collaborative filtering.

130. On information and belief, Cloudera documentation describes Cloudera Product’s generation of recommendations as:

Quick CF Recommendation Overview

From an algorithmic standpoint, the recommendation systems we’ll talk about today are considered in the k-nearest neighbor family of problems (another type would be a SVD-based recommender). We want to predict the estimated preference of a user towards an item they have never seen before. We also want to generate a ranked (by preference score) list of items the user might be most interested in. Two well-known styles of recommendation algorithms are item-based recommenders and user-based recommenders. Both types rely on the concept of a similarity function/metric (ex: Euclidean distance, log likelihood), whether it is for users or items.

⁶⁴ Class InverseUserFrequency, CLOUDERA WEBSITE, <http://archive.cloudera.com/cdh4/cdh/4/mahout-0.7-cdh4.3.2/mahout-core/org/apache/mahout/cf/taste/impl/transforms/InverseUserFrequency.html> (last visited June 14, 2015).

Josh Patterson, *Recommendation with Apache Mahout in CDH3*, CLOUDERA BLOG, November 30, 2011, <http://blog.cloudera.com/blog/2011/11/recommendation-with-apache-mahout-in-cdh3/>.

131. On information and belief, the Cloudera Products enable the calculation of a similarity value based on a test statistic for a first and second user.

132. On information and belief, one or more of the Cloudera Products enable collaborative filtering using multivariate calculations to determine a recommendation.

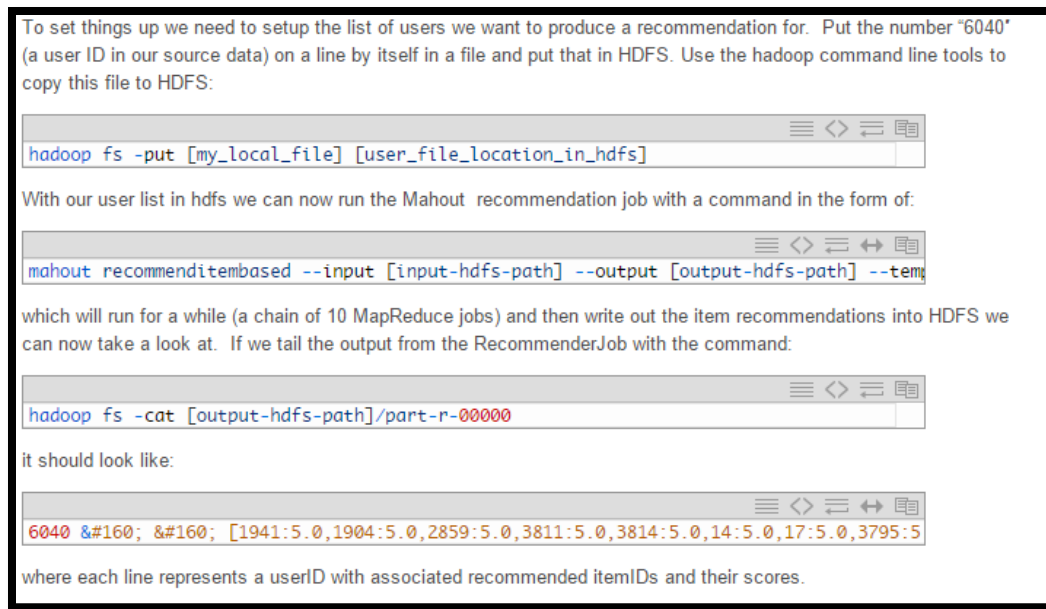
133. On information and belief, the Cloudera Products generate similarity calculations by taking into account the total number of items the users have in common.

134. On information and belief, the Cloudera Products are predictive analytics tools that enable K-means clustering analysis. K-means clustering identifies groups of similar data values in large segments of stored data.

135. On information and belief, the Cloudera Products can generate real time predictions and recommendations using a collaborative filtering engine that analyzes user interactions and ratings.

136. On information and belief, the Cloudera Products take into account the ratings distribution in recommending products and/or content.

137. On information and belief, one or more of the Cloudera Products generate recommendations as shown below:



Josh Patterson, *Recommendation with Apache Mahout in CDH3*, CLOUDERA BLOG, November 30, 2011, <http://blog.cloudera.com/blog/2011/11/recommendation-with-apache-mahout-in-cdh3/>.

138. On information and belief, the Cloudera Products use user-based matching to determine matches between a first and second user. When a first user inputs ratings data that ratings data is compared against the ratings data of other users.

139. On information and belief, the Cloudera Products compare user ratings for a common item to recommend a new item to a user.

140. On information and belief, Cloudera states in its documentation that the recommendation engine generates average ratings for content and/or products to identify unusual similarity. This identification of unusual similarity is identified in the below excerpt from Cloudera's documentation.

[w]hich boosts preference values for items for which few users have expressed a preference, and reduces preference values for items for which many users have expressed a preference. The idea is that these "rare" items are more useful in deciding how similar two users' tastes are, and so should be emphasized in other calculations.⁶⁵

⁶⁵ Class InverseUserFrequency, CLOUDERA WEBSITE, <http://archive.cloudera.com/cdh4/cdh/4/mahout-0.7-cdh4.3.2/mahout-core/org/apache/mahout/cf/taste/impl/transforms/InverseUserFrequency.html> (last visited June 14, 2015).

141. On information and belief, the Cludera Products enable creation of a factorization of a ratings matrix.

142. On information and belief, the Cludera Products enable recommendation strategies including comparing users' purchasing history against prior users of a website.

143. On information and belief, the Cludera Products includes algorithms that use averaging to improve predictive accuracy.

144. On information and belief, the Cludera Products enable the use of "k-means" to generate recommendations of products and/or content.

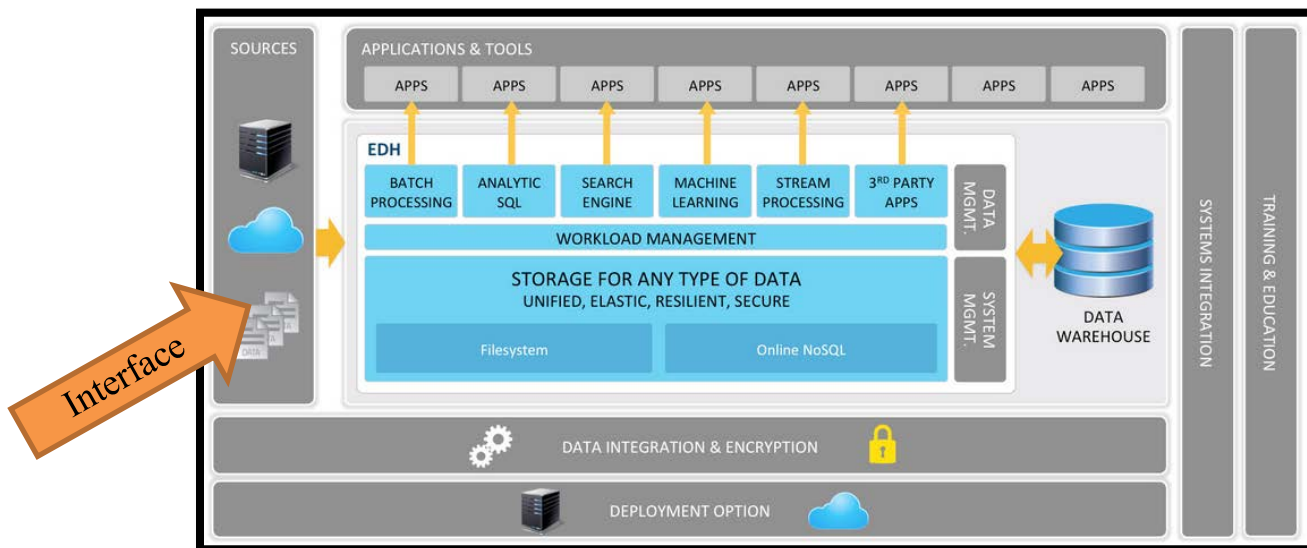
145. On information and belief, the Cludera Products enable the identification of recommended products and/or content based on linking products to users' browsing and/or purchase history.

146. On information and belief, the Cludera Products incorporate K-Nearest Neighbor and/or Naïve Bayes algorithms.

147. On information and belief, the Cludera Products generate recommendation data using an "average user," whose ratings are the average of all users' ratings.

148. On information and belief, the Cludera Products incorporate an "average user" value to improve the confidence level of recommendations.

149. On information and belief, the Cludera Products have an interface for receiving ratings data as shown in the below diagram.



Tim Stevens, *Cloudera Connect, the Blueprint to an Information-Driven Enterprise*, VISION BLOG (February 13, 2014) (slide showing (orange arrow) Cloudera Products take in ratings information (e.g., data that feeds the recommendation model)).

150. On information and belief, the Cloudera Products enable the collection of ratings data using an “information-gathering module.”

151. On information and belief, the Cloudera Products generate a numerical value as part of creating a recommendation of a product and/or content.

152. On information and belief, one or more of the Cloudera Products generate recommendations based on analyzing the entire population of users.

153. On information and belief, the Cloudera Products use algorithmic approaches to generate recommendations and preference data.

154. On information and belief, the Cloudera Products transform data associated with a user to provide product and/or content recommendations.

155. On information and belief, Cloudera has directly infringed and continues to directly infringe the ‘282 patent by, among other things, making, using, offering for sale, and/or selling collaborative filtering products and services, including but not limited to, the Cloudera Products, which include infringing content and/product recommendation technologies. Such products and/or services include, by way of example and without limitation, Cloudera Enterprise, which is covered by one or more claims of the ‘282 patent, including but not limited to claims 19 and 25.

156. By making, using, testing, offering for sale, and/or selling collaborative filtering products and services, including but not limited to the Cloudera Products, Cloudera has injured Fellowship Filtering and is liable to Fellowship Filtering for directly infringing one or more claims of the ‘282 patent, including at least claims 19 and 25, pursuant to 35 U.S.C. § 271(a).

157. On information and belief, Cloudera also infringes indirectly the ‘282 patent by active inducement under 35 U.S.C. § 271(b).

158. On information and belief, Cloudera had knowledge of the ‘282 patent since at least service of this Complaint or shortly thereafter, and on information and belief, Cloudera knew of the ‘282 patent and knew of its infringement, including by way of this lawsuit.

159. On information and belief, Cloudera intended to induce patent infringement by third-party customers and users of the Cloudera Products and had knowledge that the inducing acts would cause infringement or was willfully blind to the possibility that its inducing acts would cause infringement. Cloudera specifically intended and was aware that the normal and customary use of the accused products would infringe the '282 patent. Cloudera performed the acts that constitute induced infringement, and would induce actual infringement, with the knowledge of the '282 patent and with the knowledge, or willful blindness to the probability, that the induced acts would constitute infringement. For example, Cloudera provides the Cloudera Products that have the capability of operating in a manner that infringe one or more of the claims of the '282 patent, including at least claims 19 and 25, and Cloudera further provides documentation and training materials that cause customers and end users of the Cloudera Products to utilize the products in a manner that directly infringe one or more claims of the '282 patent. By providing instruction and training to customers and end-users on how to use the Cloudera Products in a manner that directly infringes one or more claims of the '282 patent, including at least claims 19 and 25, Cloudera specifically intended to induce infringement of the '282 patent. On information and belief, Cloudera engaged in such inducement to promote the sales of the Cloudera Products, *e.g.*, through Cloudera's user manuals, product support, marketing materials, and training materials to actively induce the users of the accused products to infringe the '282 patent. Accordingly, Cloudera has induced and continues to induce users of the accused products to use the accused products in their ordinary and customary way to infringe the '282 patent, knowing that such use constitutes infringement of the '282 patent.

160. To the extent applicable, the requirements of 35 U.S.C. § 287(a) have been met with respect to the '282 patent.

161. As a result of Cloudera's infringement of the '282 patent, Fellowship Filtering has suffered monetary damages in an amount adequate to compensate for Cloudera's infringement, but in no event less than a reasonable royalty for the use made of the invention by Cloudera together with interest and costs as fixed by the Court.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff Fellowship Filtering respectfully requests that this Court enter:

- A. A judgment in favor of Plaintiff Fellowship Filtering that Cloudera has infringed, either literally and/or under the doctrine of equivalents, the '282 patent;
- B. An award of damages resulting from Cloudera's acts of infringement in accordance with 35 U.S.C. § 284;
- C. A judgment and order requiring Cloudera to provide accountings and to pay supplemental damages to Fellowship Filtering, including, without limitation, prejudgment and post-judgment interest; and
- D. Any and all other relief to which Fellowship Filtering may show itself to be entitled.

JURY TRIAL DEMANDED

Pursuant to Rule 38 of the Federal Rules of Civil Procedure, Fellowship Filtering requests a trial by jury of any issues so triable by right.

Dated: June 19, 2015

Respectfully submitted,

/s/ Elizabeth L. DeRieux
Elizabeth L. DeRieux (TX Bar No. 05770585)
D. Jeffrey Rambin (TX Bar No. 00791478)
CAPSHAW DERIEUX, LLP
114 E. Commerce Ave.
Gladewater, Texas 75647
Telephone: 903-236-9800
Facsimile: 903-236-8787
E-mail: ederieux@capshawlaw.com
E-mail: jrambin@capshawlaw.com

OF COUNSEL:

Matt Olavi (CA SB No. 265945)
Brian J. Dunne (CA SB No. 275689)
OLAVI & DUNNE LLP
816 Congress Ave., Ste. 1620
Austin, Texas 78701
Telephone: 512-717-4485
Facsimile: 512-717-4495
E-mail: molavi@olavidunne.com
E-mail: bdunne@olavidunne.com

Dorian S. Berger (CA SB No. 264424)
Daniel P. Hipskind (CA SB No. 266763)
OLAVI & DUNNE LLP
1880 Century Park East, Ste. 815
Los Angeles, CA 90067
Telephone: 213-516-7900
Facsimile: 213-516-7910
E-mail: dberger@olavidunne.com
E-mail: dhipskind@olavidunne.com

*Attorneys for Fellowship Filtering
Technologies LLC*