

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

UIPATH, INC.,

Petitioner,

v.

RULE 14 LLC.,

Patent Owner.

Case No. IPR2025-00623
U.S. Patent No. 10,108,679

DECLARATION OF DR. JAIDEEP SRIVASTAVA, PH.D.

Table of Contents

I.	INTRODUCTION	1
II.	EXPERIENCE AND QUALIFICATIONS	1
III.	MATERIALS CONSIDERED	4
IV.	LEGAL STANDARDS	7
	A. Written Description Support	7
	B. Anticipation	8
	C. Obviousness.....	9
V.	LEVEL OF ORDINARY SKILL IN THE ART	11
VI.	THE SPECIFICATION AND CLAIMS OF THE '679 PATENT	12
VII.	THE PROSECUTION HISTORY OF THE '679 PATENT.....	13
VIII.	THE EFFECTIVE FILING DATE OF THE CLAIMS OF THE '679 PATENT IS NO EARLIER THAN JUNE 26, 2013.....	14
IX.	CLAIM CONSTRUCTION.....	16
	A. “expand[ing]search terms of the query” [Claim 6]	17
	B. “extractors” [Claim 8]	17
	C. “[establishing/establish] a... communication channel” [Claim 1, 11, 16] 17	
	D. “generat[e/ing], via an entity, a query based at least in part on a topic of interest” [Claims 1, 11 and 16]	18
	E. “entity” [Claims 1, 10-11, 16]	18
	F. Terms Identified By Petitioner As Indefinite In The Litigation	18
X.	PRIOR ART RELIED ON IN PETITION	21
XI.	SUMMARY OF GROUNDS FOR UNPATENTABILITY	29
XII.	ALL OF THE CHALLENGED CLAIMS ARE OBVIOUS.....	30
	A. Ground 1 - Claims 1, 2, 8-12, 15-16 and 19-20 Are Obvious Over Safford, Chen and Laribi	30
	1. Claim 1	30
	2. Claims 11 and 16.....	55

3. Claims 2 and 12 - “visualizing the extracted data”	56
4. Claim 8 - “the monitoring comprises deploying extractors to the data source, a number of extractors based at least in part on a traffic volume of the data source”	56
5. Claims 9, 15 and 20 - “in which the two-way communication channel comprises at least one of ... social media”	57
6. Claim 10	57
7. Claim 19 - “the plurality of data sources are predefined”	59
B. Ground 2 - Claims 3, 13 and 17 Are Obvious Over Safford, Chen and Laribi.....	59
C. Ground 3 - Claims 4, 14 and 18 Obvious Over Safford, Chen and Laribi In View Of Seibel	61
D. Ground 4 - Claims 6-7 Are Obvious Over Safford, Chen, Laribi, Schultz And Gravano	63
E. Ground 5 - Claim 5 Is Obvious Over Safford, Chen and Laribi, In View Of Schultz and Moulinier	68
F. Ground 6 - Claims 1, 2, 8-12, 15-16 and 19-20 are Obvious Over Safford, Chen and Masters	69
1. Claim 1	69
2. Claims 2, 11, 12 and 16	81
3. Claim 8	81
4. Claims 9, 10, 15, 19 and 20	82
G. Ground 7 - Claims 3, 13 and 17 Are Obvious Over Safford, Chen and Masters.....	83
H. Ground 8 - Claims 4, 14 and 18 Obvious Over Safford, Chen and Masters In View Of Seibel	83
I. Ground 9 - Claims 6-7 Are Obvious Over Safford, Chen, Masters, Schultz and Gravano.....	84
J. Ground 10 - Claim 5 Is Obvious Over Safford, Chen and Masters, In View Of Schultz and Moulinier	84
K. Ground 11 - Claims 1, 2, 8-12, 15-16 and 19-20 Are Obvious Over Heath, Randall and Laribi.....	85
1. Claim 1	85

2. Claims 11 and 16.....	97
3. Claims 2 and 12.....	97
4. Claim 8 - “the monitoring comprises deploying extractors to the data source, a number of extractors based at least in part on a traffic volume of the data source”	98
5. Claims 9, 15 and 20.....	98
6. Claim 10.....	99
7. Claim 19.....	99
L. Ground 12 - Claims 3, 13 and 17 Are Obvious Over Heath, Randall, Laribi and Chen	100
M. Ground 13 - Claims 4, 14 and 18 Obvious Over Heath, Randall and Laribi In View Of Seibel	100
N. Ground 14 - Claims 6-7 Are Obvious Over Heath, Randall, Laribi, Schultz and Gravano.....	101
O. Ground 15 - Claim 5 Is Obvious Over Heath, Randall and Laribi, In View Of Schultz and Moulinier	102

I. INTRODUCTION

1. My name is Dr. Jaideep Srivastava. I have been retained as an expert witness on behalf of Petitioner to provide expert opinions on the patentability of U.S. Patent No. 10,108,679 (the “’679 patent”) (Ex. 1001).

2. I am being compensated at my normal rate, plus reimbursement for expenses, for my analysis. My compensation does not depend on the content of my opinions or the outcome of this proceeding.

II. EXPERIENCE AND QUALIFICATIONS

3. My qualifications are set forth in my curriculum vitae, a copy of which is submitted as Ex. 1005.

4. I received my Bachelor’s degree from the Indian Institute of Technology (IIT), India (1983), and my Master’s (1985) and PhD (1988) degrees from the University of California – Berkeley. All of my degrees are in Computer Science.

5. I am currently a Professor of Computer Science and Engineering at the University of Minnesota, a position I have held since 1988.

6. I have more than forty (40) years of technical expertise in the field of computer science and engineering, including experience with databases, optimization techniques, applications of graphs and database system design and architecture.

7. I have taught a range of database courses at the University of Minnesota, including core database courses like *Principles of Database Systems*, *Database System Implementation*, *Introduction to Database Research*, and specialized ones like *Introduction to Social Network Analysis* – which have strong database components.

8. I have supervised forty-six (46) doctoral dissertations and seventy (70) Masters theses/projects in the field of Computer Science & Engineering. Sixteen (16) of my PhD advisees themselves are professors in various universities across the world, while many of the rest are in some of the top technology companies. I have also mentored a number of university undergraduates and high school students, many of which have gone on to careers in the STEM disciplines.

9. I have authored or co-authored over four hundred sixty (460) papers in peer-reviewed journals and conferences in the field of Computer Science & Engineering.

10. I am a co-inventor on six (6) United States patents in the field of Computer Science & Engineering.

11. My research in the field of Computer Science & Engineering has been funded by various federal agencies including the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), the Defense Advanced Research Projects Agency (DARPA), the Intelligence Advanced

Research Projects Agency (IARPA), the U.S. Army, and the US Air Force – via multiple, peer-reviewed grants. My research has also been supported by the State of Minnesota and various private organizations.

12. I have worked with a number of commercial database systems while serving as a Data Mining Architect at Amazon from December 1999 through June 2000, and the Director of Data Analytics at Yodlee Inc. from July 2000 through June 2001. In my role as Director of Data Analytics at Yodlee, Inc., I led the creation of an analytics system from scratch, and was responsible for various technical decisions related to it.

13. I have advised a number of companies on database technologies, including Apertus, Honeywell, Eaton and United Technologies, among others.

14. I was an advisor to the State of Minnesota from 2006 through 2008 and the Government of India from 2011 through the present, in the use of database and other information technologies for more effective governance.

15. In 2010, I co-founded an analytics company, and was responsible for overseeing the end-to-end development of the first version of the system. I was responsible for the architecture and other technical aspects of the system, and leading the team that did the programming and development. A SQL database was at the core of the system, and I was personally responsible for the design, architecture, and implementation decisions around it.

16. I am a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the world's largest professional society of Electrical Engineering, Communications and Computing professionals. I have been an IEEE Distinguished Visitor, and received the PAKDD's Lifetime Achievement Award for my contributions to the field of data mining. I have served on various journal editorial boards and conference organizing committees – many times chairing the technical committees. I have been invited to deliver over fifty (50) keynote speeches, tutorials and other invited talks at national and international conferences.

17. I am an active researcher, with eight (8) current PhD and four (4) current MS students, working in the areas of Social Media Analytics and Health Informatics.

18. My analyses set forth in this Declaration are informed by my experience in the field of computer science, including computer architecture, operating systems, distributed systems, database systems, data mining and machine learning. Based on my above-described experience, I believe that I am considered to be an expert in these fields.

III. MATERIALS CONSIDERED

19. In preparing this Declaration, I have reviewed the '679 patent and file history. I also have considered the accompanying Petition, and the following documents cited in this Declaration:

Exhibit No.	Description
Ex. 1001	U.S. Patent No. 10,108,679
Ex. 1002	File History of U.S. Patent Appl. 14/960,309, filed Dec. 4, 2015
Ex. 1003	Provisional Appl. No. 61/679,024
Ex. 1004	Expert Declaration of Jaideep Srivastava, Ph.D.
Ex. 1005	<i>Curriculum Vitae</i> of Jaideep Srivastava, Ph.D.
Ex. 1006	U.S. Patent No. 9,288,123 (Safford)
Ex. 1007	U.S. Patent No. 8,909,771 (Heath)
Ex. 1008	U.S. Patent No. 9,769,085 (Laribi)
Ex. 1009	Provisional Application No. 61/643,114, filed May 4, 2012.
Ex. 1010	U.S. Patent No. 7,051,098 (Masters)
Ex. 1011	U.S. Patent No. 8,010,524 (Chen)
Ex. 1012	U.S. Patent No. 5,721,902 (Schultz)
Ex. 1013	U.S. Patent No. 7,315,861 (Seibel)
Ex. 1014	U.S. Patent No. 9,418,139 (Moulinier)
Ex. 1015	Gravano et al., <i>Precision and Recall of GLOSS Estimators for Database Discovery</i> , Proceedings Of The International Conference On Parallel And Distributed Information Systems, pp. 103-106, 1994 (Gravano)
Ex. 1016	Infringement Contentions of Patent Owner
Ex. 1017	ISBN Lookup results for ISBN No. 0-8186-6400-2 for Text, <i>Proceedings of 3rd International Conference on Parallel and Distributed Information Systems</i>
Ex. 1018	Text, <i>Proceedings of 3rd International Conference on Parallel and Distributed Information Systems</i> , 28-30 Sept. 1994, selected pages (ISDN No. 0-8186-6400-2)
Ex. 1019	International Search Report (ISR) for International Application No. PCT/US98/18844, International Application No. WO1999014691A1, dated December 8, 1998
Ex. 1020	Bibliographic data from Espacenet for XP 002085758, for <i>Precision and Recall of GLOSS Estimators for Database Discovery</i>

Exhibit No.	Description
Ex. 1021	Webpage from IEEE Explore Database, for <i>Precision and Recall of GLOSS Estimators for Database Discovery</i>
Ex. 1022	United States Reissue Patent No. RE45752 and file history
Ex. 1023	U.S. Patent No. 7,725,452 (Randall)
Ex. 1024	Declaration of Shari Hall White - Exhibit A: IDS from United States Reissue Patent No. RE45752 - Exhibit B: <i>Precision and Recall of GLOSS Estimators for Database Discovery</i>, submitted to USPTO
Ex. 1025	Declaration of Desiree L. Wolf - Exhibit A: Web page, https://ieeexplore.ieee.org/document/331726 - Exhibit B: Web page, https://ieeexplore.ieee.org/xpl/conhome/997/proceeding - Exhibit C: Web page, https://ieeexplore.ieee.org/xpl/conhome/997/proceeding?isnumber=7843&sortType=vol-only-seq&pageNumber=2 - Exhibit D: Web page, https://isbnsearch.org/ - Exhibit E: Web page, https://isbnsearch.org/isbn/0818664002 - Exhibit F: Web page, https://worldwide.espacenet.com/?locale=EN_ep - Exhibit G: https://worldwide.espacenet.com/searchResults?ST=singleline&locale=en_EP&submitted=true&DB=&query=XP002085758 - Exhibit H: Web page, https://worldwide.espacenet.com/nplReferenceDetails/biblio?CC=XP&NR=002085758&KC=&FT=E&DB=&locale=en_EP&ND=1&II=0
Ex. 1026	Declaration of Daniel H. Golub

Exhibit No.	Description
	- Exhibit A: Copy of <i>Precision and Recall of GLOSS Estimators for Database Discovery</i>, purchased from https://ieeexplore.ieee.org/document/331726 - Exhibit B: Purchase email receipt for text, <i>Proceedings of 3rd International Conference on Parallel and Distributed Information Systems</i>, 28-30 Sept. 1994
Ex. 1027	U.S. Patent Publ. No. 2010/0246797
Ex. 1028	U.S. Patent Publ. No. 2006/0253427
Ex. 1029	U.S. Patent No. 8,448,084
Ex. 1030	International Publ. No. WO 2011/139643
Ex. 1031	U.S. Patent Publ. No. 2013/0014258
Ex. 1032	Joint Claim Construction Chart

In formulating my opinions, I have further relied upon my extensive experience.

IV. LEGAL STANDARDS

A. Written Description Support

20. I understand that a claim is not entitled to the benefit of an earlier filed application unless the subject matter of the claim is disclosed in the earlier filed application in a manner that reasonably conveys to those skilled in the art that the inventor had possession of and actually invented the claimed subject matter. I understand that adequate written description requires that one skilled in the art, reading the original disclosure, be able to immediately discern all limitations in the claim.

21. I understand that the test for adequate written description support requires an objective inquiry into the four corners of the specification from the perspective of a person of ordinary skill in the art. I understand that the specification need not recite the claimed invention *in haec verba*. I also understand, however, that a description that merely renders the invention obvious does not satisfy the written description requirement. In other words, entitlement to a filing date does not extend to subject matter which is not disclosed but would be obvious over what is expressly disclosed.

B. Anticipation

22. It is my understanding that the claims of a patent are anticipated by a prior art reference if each and every element of the claim is found either explicitly or inherently in a single prior art reference or system. I understand that inherency requires a showing that the missing descriptive matter in the claim is necessarily or implicitly present in the allegedly anticipating reference, and that it would have been so recognized by a person of ordinary skill in the art (POSITA) at the time. In addition, I understand that an enabling disclosure is one that allows a POSITA to make the invention without undue experimentation. Although anticipation typically involves the analysis of a single prior art reference, I understand that additional references may be used to show that the primary reference has an enabling

disclosure, to explain the meaning of a term used in the primary reference, and/or to show that a characteristic is inherent in the primary reference.

C. Obviousness

23. I understand that a claim is unpatentable if the differences between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious to a POSITA at the time the claimed subject matter was made. I understand that a patent claim may be obvious to a POSITA in view of the prior art teachings of a single reference or a combination of references. I understand that a claim is obvious based on a combination of references where a POSITA would have been motivated to make such a combination, and would have had a reasonable expectation of success in arriving at the claimed subject matter.

24. To assess obviousness, I understand that I am to consider the scope and content of the prior art, the differences between the prior art and the claim, the level of ordinary skill in the art, and any secondary considerations to the extent they exist.

25. It is also my understanding that there are principles that may be used as further guidance in obviousness analysis (especially when considering combinations of references), which include considering whether:

- the claimed subject matter is simply a combination of prior art elements according to known methods to yield predictable results;

- the claimed subject matter is a simple substitution of one known element for another to obtain predictable results;
- the claimed subject matter uses known techniques to improve similar devices or methods in the same way;
- the claimed subject matter applies a known technique to a known device or method to yield predictable results;
- the claimed subject matter would have been “obvious to try” choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- there is known work in one field of endeavor that may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to a POSITA;
- there existed at the time of conception and reduction to practice a known problem for which there was an obvious solution encompassed by the patent’s claims; and
- there is some teaching, suggestion, or motivation in the prior art that would have led a POSITA to modify the prior art reference or to

combine prior art reference teachings to arrive at the claimed subject matter.

26. I understand that there are secondary considerations that may tend to show that a claimed invention would not have been obvious to a POSITA. These include, for example, commercial success, long-felt but unsolved needs, failure of others, and unexpected results. I further understand that there must be a relationship, or nexus, between any secondary considerations and the novelty of the claimed invention.

V. LEVEL OF ORDINARY SKILL IN THE ART

27. I am informed and understand that various factors can be considered in determining the skill and experience of a POSITA. I am informed and understand that those factors include: (1) the educational level of the inventors; (2) the type of problems encountered in the art; (3) prior art solutions to those problems; (4) the rapidity with which innovations are made; (5) sophistication of the technology; and (6) education level of active workers in the field.

28. A POSITA as of June 26, 2013 would have had a working knowledge of techniques for querying and monitoring data sources as pertinent to the '679 Patent. A POSITA would have had a bachelor's degree in electrical engineering, computer science, or equivalent training, and approximately two years of work experience in software development with a basic knowledge of database searching

techniques. My definition of the levels of education and experience of a POSITA would be the same as of August 2, 2012 (the filing date of Provisional Application No. 61/679,024 (Ex. 1002)) and June 26, 2013 (the filing date of Application No. 13/928,313.)

VI. THE SPECIFICATION AND CLAIMS OF THE '679 PATENT

29. The '679 Patent describes “a data mining system” in which “a query” is used for searching a plurality of data sources including, for example, social media postings. Ex. 1001, 4:16-20 and 4:31-58. Next, “the results of the query are reviewed.” *Id.*, 4:59. A given data source is “selected to be monitored” for updates if there is a “correlation” between the query and the results provided by such data source, *e.g.*, if the data source provides “relevant” or “accurate” search results. *Id.*, 4:59-61 and 6:57-64. The selected data sources are periodically monitored for updates that match the original query or a refined query, and when there is a match, “a data extractor extracts the data” which is analyzed to identify a “sentiment” -- such as the sentiment expressed by a customer about a company’s product or service on social media. *Id.*, 4:62-5:2, 7:49-52 and 15:38-41. In one example, an “appropriate response” (*e.g.*, customer outreach) follows. *Id.*, 18:41-46.

30. The '679 Patent explains that the extractors (used for ongoing monitoring) “may be bots or crawlers” deployed on social media. *Id.*, 7:25-30. In the '679 Patent, the number of bots or crawlers “deployed for a data source increases

when activity on the data source increases” which “allows for on-demand scaling of resources and deployment of multiple extractors that may be processed in parallel. *Id.*, 7:25-30, 14:10-12 and 14:54-65.

VII. THE PROSECUTION HISTORY OF THE '679 PATENT

31. The '679 Patent matured from Appl. No. 14/960,309, filed December 4, 2015. In response to three office actions, the Applicant made several claim amendments and arguments all of which failed to result in allowability of the application. See, Ex. 1002, 648-663 (first office action rejecting claims over Gonzalez); 617-625 (response to first office with claim amendments); 505-522 (final office action rejecting claims over Gonzalez and Aggarwal); 462-475 (response following final office action with further claim amendments); 366-386 (further office action maintaining rejection over Gonzalez and Aggarwal); 349-362 (response to further office action with still further claim amendments); and 241-263 (further office action rejecting claims over Gonzalez and Herring).

32. The Applicant responded to the last office action (Ex. 1002, 209-238) by amending claim 1 to include elements 1.6 through 1.9 (identified in Section X(A)(1)(e)-(h) below), which relate to the determination of a data extraction rate, and the dynamic adjustment of the number of parallel processors performing data analysis. A Notice of Allowance was then issued.

33. The prosecution history demonstrates that all of the elements of Claim 1 were present in the references applied by the Examiner, except for, arguably, the determination of a data extraction rate, and the dynamic adjustment of the number of parallel processors performing data analysis.

VIII. THE EFFECTIVE FILING DATE OF THE CLAIMS OF THE ‘679 PATENT IS NO EARLIER THAN JUNE 26, 2013

34. The ‘679 Patent is a continuation-in-part of Appl. No. 13/928,313, filed June 26, 2013, which in turn claimed priority to Provisional Appl. No. 61/679,024 filed August 2, 2012. Provisional Appl. No. 61/679,024 was a one-page marketing document (below) that fails to support any of the Challenged Claims:

RULE14™ LOSS PREVENTION SERVICES

Introducing a new and innovative way to prevent loss combining extensive customer intelligence with an automated system for you and your customers to communicate more effectively with one another.

Rule14™ LPS uses four integrated modules:

-  Listen™ - finds the events and topics that concern your customers which are gathered from a myriad of channels where customers are discussing your service
-  Pulse™ - takes these critical topics and generates a dashboard of actionable intelligence so you can understand your customers and where they're headed
-  Reachout™ - uses this actionable intelligence to communicate effectively and directly with each one of your customers to improve satisfaction and recoveries
-  Speakout™ - to easily generate and garner feedback from any customer on any topic

Property of Rule14 LLC | Proprietary and Confidential



For example, the Provisional Application makes no mention of any of the following limitations recited in the independent claims of the '679 Patent: "selecting, by the entity, a data source from the plurality of data sources for monitoring based on a correlation between the data source and the topic of interest, the correlation determined based on results of the executed query"; "monitoring, based on a set schedule, the data source for matches to the query to extract data from the data source when at least an update to stored data matches the query, newly added data matches the query, or a combination thereof"; "extracting data from the data source when at least the update to the stored data matches the query, the newly added data matches

the query, or the combination thereof"; "determining an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period"; determining a first processing rate for processing the extracted data with a number of parallel processors, the first processing rate indicating an amount of extracted data that is processed over a second time period"; "dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate to obtain a second processing rate that is greater than the first processing rate." See, Ex. 1003, at 14. Thus, the effective filing date of the claims of the '679 Patent is no earlier than June 26, 2013.¹ The references relied on herein and in the Petition predate June 26, 2013.

IX. CLAIM CONSTRUCTION

35. I understand that claim terms should normally be given their ordinary and accustomed meaning as would have been understood by a POSITA as of the priority date in view of the specification and prosecution history. In formulating my opinions, and as set forth below with respect to specific terms, I have, to the extent ones exists, applied the ordinary and accustomed meaning as would have been understood by a POSITA as of the priority date in view of the specification and prosecution history to the claims of the '679 Patent.

¹ I express no opinion as to whether the claims of the '679 Patent are supported by Appl. No. 13/928,313, filed June 26, 2013.

A. “expand[ing]search terms of the query” [Claim 6]

36. This term should be interpreted as “applying a lexicon to add additional terms to the search query.” Ex. 1001, ‘679 Patent, 4:21-30; 6:37-49; and 9:43-59. Patent Owner contends that no construction is necessary. Ex. 1032, at 3. As explained below, the prior art discloses a lexicographical analysis such as a semantic network that expands search terms of a query.

B. “extractors” [Claim 8]

37. This term should be interpreted as “applications that monitor and extract data from user-specified locations or sources of interest (e.g., data source).” Ex. 1001, ‘679 Patent, 4:65-5:2; 7:9-35. The user-specified locations or sources may be the data sources to which the original query is applied. The Patent Owner contends that no construction is necessary. Ex. 1032, at 3. As explained below, the prior art discloses applications that monitor and extract data from user-specified sources of interest.

C. “[establishing/establish] a... communication channel” [Claim 1, 11, 16]

38. This term should be interpreted as “set[ting] up a communication channel for later communication with a user or user device.” Ex. 1001, 5:6-21; 8:29-54. Patent Owner contends that no construction is necessary. Ex. 1032, at 3. As explained below, the prior art discloses setting up a communication channel for later communication with a user.

D. “generat[e/ing], via an entity, a query based at least in part on a topic of interest” [Claims 1, 11 and 16]

39. The meaning of this term as used in the '679 Patent requires that the query be generated via an entity, meaning an organization, which would include a human user acting on behalf of the entity. Ex. 1001, Claim 10 (referring to "entity user"). In the Litigation, the parties dispute whether the claim *also* covers queries generated by a computer and, if so, whether the claim is indefinite when interpreted that way. Ex. 1032, at 4. As explained below, the prior art discloses or renders obvious queries generated by an entity (which includes a human user acting on behalf of an entity).

E. “entity” [Claims 1, 10-11, 16]

40. This term should be interpreted as “a separate organization such as businesses or police, not any user.” Ex. 1001, ‘679 Patent, at 6:13-18; Ex. 1002, 219-221, 223, 463. For clarity, a human user can act on behalf of an "entity" but other users (not associated with the entity) are outside the scope of an "entity." Patent Owner contends that no construction of this term is necessary. Ex. 1032, at 4. The Patent Owner’s Infringement Contentions are consistent with Petitioner’s proposed construction. Ex. 1016, 15-32. As explained below, the prior art discloses or renders obvious operation by a separate organization such as a business.

F. Terms Identified By Petitioner As Indefinite In The Litigation

41. The outer scope of certain claim limitations (discussed below) is indefinite for the purpose of determining infringement. However, such

indefiniteness does not render it logically impossible to determine that the prior art satisfies the limitations because each of the limitations covers at least some subject matter which is met by the prior art.

42. By way of analogy, while the precise boundary line dividing New York and New Jersey is disputed, it is undisputed that the City of Trenton (many miles south of the disputed border) is within New Jersey. Such is the case for each term discussed below -- while the outer contours of what is covered by the scope of the claims is subject to dispute, such indefiniteness does not render it logically impossible to assess validity in view of the prior art because a POSITA would understand that there is subject matter (analogous to Trenton) within the disputed contours that is met by the prior art.

1. “determine[e/ing] an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period” (Claims 1, 11 and 16)

43. The outer contours of this limitation are unclear because the ‘679 Patent does not, for example, clearly delineate between metrics that do and do not constitute an extraction rate. For example, the ‘679 Patent does not clarify if the ratio of the quantity of data extracted over a given time divided by the volume of traffic analyzed during the same time is an “extraction rate.” However, while the outer contours are unclear, similar to the city of Trenton in the analogy above, this term covers

determining a quantity of data output by the extraction process over a period of time.

Patent Owner agrees. Ex. 1016, 105-110, 453, and 728.

2. **“determine[e/ing] a first processing rate for processing the extracted data with a number of parallel processors, the first processing rate indicating an amount of extracted data that is processed over a second time period” (Claims 1, 11 and 16)**

44. The outer contours of this limitation are unclear because the '679 Patent does not, for example, clearly delineate between the types of data “processing” used and not used for determining “a first processing rate.” For example, it is not clear whether processing the extracted data to identify duplicative entries would constitute “processing the extracted data” for purposes of determining a first processing rate. However, while the outer contours are unclear, similar to the city of Trenton in the analogy above, this term covers determining a quantity of extracted data analyzed over a second period of time to determine a sentiment. Patent Owner agrees. Ex. 1016, 110-115, and 134.

3. **“[to] dynamically adjust the number of parallel processors for analyzing the extracted data based on the extraction rate to obtain a second processing rate that is greater than the first processing rate” (Claims 1, 11 and 16)**

45. The outer contours of this limitation are unclear because the '679 Patent does not, for example, clearly delineate the specifics of how any “parallel processors” would be adjusted, or clearly delineate between arrangements that are

and are not within the scope of “parallel processors.” However, while the outer contours are unclear, similar to the city of Trenton in the analogy above, this term covers dynamically adjusting “parallel processors” that are distinct processors operating in parallel. Ex. 1001, ‘679 Patent, at 25:61–64; 27:16–19; 28:24–27.

4. “accuracy threshold[s]” (Claim 6)

46. The outer contours of this limitation are unclear because the '679 Patent does not, for example, clearly delineate between what thresholds of other metrics are within and not within the scope of “accuracy threshold[s].” For example, it is not clear whether a threshold of a metric that measures the percentage of retrieved documents from the database that had one or more keywords included in the query, would constitute an “accuracy threshold[s].” While the outer contours are unclear, similar to the city of Trenton in the analogy above, this term covers a threshold of a metric that measures for a given database the percentage of retrieved documents that were actually relevant (not false positives) to a query. Ex. 1002, at 360-361.

X. PRIOR ART RELIED ON IN PETITION

47. The prior art relied on herein and in the Petition is identified below.

a. U.S. Patent No. 9,288,123 (“Safford”) discloses techniques for analyzing social media accounts associated with a company or brand. Ex. 1006, at 3:14-18.

b. U.S. Patent No. 8,909,771 (Heath) discloses that collects and analyzes online social media online communications for brand monitoring. Ex. 1007, 1:30-44.

c. U.S. Patent No. 9,769,085 (Laribi) discloses a method for automatically scaling a number of instances of an application based on metrics and a threshold. Ex. 1008, 1:31-46, 1:51-2:2 and 2:7-10. Laribi was filed on May 3, 2013, which predates the earliest possible effective filing date of the claims of the '679 Patent of June 13, 2013. Moreover, Laribi claimed priority to Provisional Application 61/643,114, filed May 2, 2012 (Ex. 1010), and all of the aspects of Laribi relied on herein are supported in Laribi's Provisional Application.

d. U.S. Patent No. 7,051,098 (Masters) discloses quality of service (QoS) managers for “recommending that applications be either scaled up, scaled down, or moved in order to maintain acceptable performance.” Ex. 1010, 18:45-47 and 53-58.

e. U.S. Patent 8,010,524 (Chen) taught the use of refined queries in order to eliminate “irrelevant information.” Ex. 1011, 6:39-41.

f. U.S. Patent No. 5,721,902 (Schultz) taught natural language search queries. Ex. 1012, 4:8-9.

g. U.S. Patent No. 7,315,861 (Seibel) taught a text mining system for business intelligence that performed text mining based on internal corporate documents “such as emails” and “customer support notes.” Ex. 1013, 9:28-29 and 5:33-35.

h. U.S. Patent No. 9,418,139 (Moulinier) taught translation of a natural language query from one language to another. Ex. 1014, 4:10-14 and 6:3-57.

i. U.S. Patent No. 7,725,452 (Randall) discloses crawlers or robots for detecting changes in online content, and the scheduling of crawling at regular intervals. Ex. 1023, 2:66-3:1, 5:50-59, 6:1.

j. Gravano et al., *Precision and Recall of GLOSS Estimators for Database Discovery* (Gravano) discloses a metric which reflects, for a given database and query, the percentage of retrieved documents from the database that were false positives. Ex. 1015, at 3. Gravano is a printed publication, publicly available before the priority date of the ‘679 Patent. Gravano was published by the IEEE in “Proceedings of 3rd International Conference on Parallel and Distributed Information Systems,” September 28-30, 1994 and catalogued under ISBN No. 0-8186-6400-2.

ISBN 9780818664007 - Proceed... x +

← → ↻ isbnsearch.org/isbn/0818664002

ISBN Search

Find a book:

0-8186-6400-2 [Search](#)



Proceedings of the Third International Conference on Parallel and Distributed Information Systems: September 28-30, 1994 Austin, Texas

ISBN-13: 9780818664007
ISBN-10: 0818664002

Authors: Intl. Conf. on Parallel; Distributed Information Systems
Binding: Paperback
Publisher: IEEE
Published: 1994-09-01

[Sell this book](#) [Buy or Rent?](#)

Best Used Prices

 **\$47.83**
+4.49 shipping [Notes](#) [Buy this book](#)

Ex. 1017; Ex. 1025 (Exs. D & E). Excerpts of the text showing a copyright date of 1994 is attached as Ex. 1018 (see Ex. 1026, Ex. B).



IEEE Computer Society Press
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1264

Copyright © 1994 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society Press, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Press Order Number 6400-02

Library of Congress Number 94-76453

IEEE Catalog Number 94TH0668-4

ISBN 0-8186-6400-2 (paper)

ISBN 0-8186-6401-0 (microfiche)

Additional copies may be ordered from:

IEEE Computer Society Press
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1264
Tel: +1-714-821-8380
Fax: +1-714-821-4641
Email: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: +1-908-981-1393
Fax: +1-908-981-9667

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM
Tel: +32-2-770-2198
Fax: +32-2-770-8505

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama
Minato-ku, Tokyo 107
JAPAN
Tel: +81-3-3408-3118
Fax: +81-3-3408-3553

Editorial production by Mary E. Kavanaugh

Cover design by Joseph Daigle

Printed in the United States of America by KNI, Inc.



The Institute of Electrical and Electronics Engineers, Inc.

An Efficient Parallel Join Algorithm Based on Hypercube-Partitioning	50
<i>H.I. Choi, B.M. Im, M.H. Kim, and Y-J. Lee</i>	
An Economic Paradigm for Query Processing and Data Migration in Mariposa	58
<i>M. Stonebraker, R. Devine, M. Kornacker, W. Litwin, A. Pfeffer, A. Sah, and C. Staelin</i>	
Industry Session 3B: Parallel Database Servers	
Chair: Lynn Streeter, <i>US West</i>	
The RISAM Storage Manager for Parallel Architectures	69
<i>A. Tulshibagwale, S. Parikh, S. Mahajan, T. Tumbay, and K.C. Pravin</i>	
[additional talks to be announced during conference]	
Parallel Sessions III —	
Session 4A: Mobile Computing and Information Exchange	
Chair: Robert Meersman, <i>University of Tilbourg</i>	
Checkpointing Distributed Applications on Mobile Computers	73
<i>A. Acharya and B.R. Badrinath</i>	
Location Management in Distributed Mobile Environments	81
<i>P. Krishna, N.H. Vaidya, and D.K. Pradhan</i>	
Distributed Selective Dissemination of Information	89
<i>T.W. Yan and H. Garcia-Molina</i>	
Synopsis — Secure Access to Data Over the Internet	99
<i>E. Bina, V. Jones, R. McCool, and M. Winslett</i>	
Synopsis — Precision and Recall of GLOSS Estimators for Database Discovery	103
<i>L. Gravano, H. Garcia-Molina, and A. Tomasic</i>	
Session 4B: Panel — Legacy Systems — the Achilles Heel of Downsizing	
Chair: Michael Stonebraker, <i>University of California — Berkeley</i>	
108	
Panelists:	
David Kessel, <i>Pacific Telesis</i>	
David Lomet, <i>DEC Cambridge Research Lab</i>	
Marie-Anne Neimat, <i>Hewlett-Packard Labs</i>	
Parallel Sessions IV —	
Session 5A: Parallel Processing and Replicated Data	
Chair: Dan Miranker, <i>University of Texas at Austin</i>	
Parsets for Parallelizing OODBMS Traversals: Implementation and Performance	111
<i>D.J. DeWitt, J.F. Naughton, J.C. Shafer, and S. Venkataraman</i>	
Query Processing for Decision Support: The SQLmp Solution	121
<i>L.S. Colby, N.L. Martin, and R.M. Wehrmeister</i>	
Divergence Caching in Client-Server Architectures	131
<i>Y. Huang, R.H. Sloan, and O. Wolfson</i>	
Session Guarantees for Weakly Consistent Replicated Data	140
<i>D.B. Terry, A.J. Demers, K. Petersen, M.J. Spreitzer, M.M. Theimer, and B.B. Welch</i>	

On December 8, 1998, Gravano was located by the European Patent Office and cited in the International Search Report (ISR) for International Application No. PCT/US98/18844:

INTERNATIONAL SEARCH REPORT		
		In. .ional Application No PCT/US 98/18844
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No.
A	GRAVANO L ET AL: "Precision and recall of GIOSS estimators for database discovery" PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON PARALLEL AND DISTRIBUTED INFORMATION SYSTEMS (CAT. NO.94TH0668-4), PROCEEDINGS OF 3RD INTERNATIONAL CONFERENCE ON PARALLEL AND DISTRIBUTED INFORMATION SYSTEMS, AUSTIN, TX, USA, 28-30 SEPT. 1994, pages 103-106, XP002085758 ISBN 0-8186-6400-2, 1994, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA see page 103, left-hand column, paragraph 1 - page 104, left-hand column, paragraph 3	1,10,15, 24,28
A	"PARCI ARCHITECTURE" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 38, no. 3, 1 March 1995, pages 13-18, XP000507963 see the whole document -----	1,10,15, 24,28

Ex. 1019, at 46. The ISR referenced Gravano by both its ISBN number (green arrow), and an “XP” number (blue arrow) used by the EPO for cataloging references.



Espacenet
Patent search

Deutsch English Français
Contact
Change country

About Espacenet Other EPO online services

Search Result list My patents list (0) Query history Settings Help

Refine search → Results → XP002085758

Bibliographic data

Description
Claims
Abstracts
Original document
Cited documents
Citing documents
INPADOC legal status
INPADOC patent family

Quick help

★

In my patents list

Previous

1/2

Next

Report data error

Print

Precision and recall of GIOSS estimators for database discovery

Page bookmark

XP002085758 - Precision and recall of GIOSS estimators for database discovery

NPL reference number:

XP 002085758

Publication date:

Author:

GRAVANO L, ET AL

Editor:

cooperative:

Publication data:

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON PARALLEL AND DISTRIBUTED INFORMATION SYSTEMS, 19940928 XX, XX

Source information:

Page(s) 103 - 106

Publisher accession number:

Patent applicant:

Publication number:

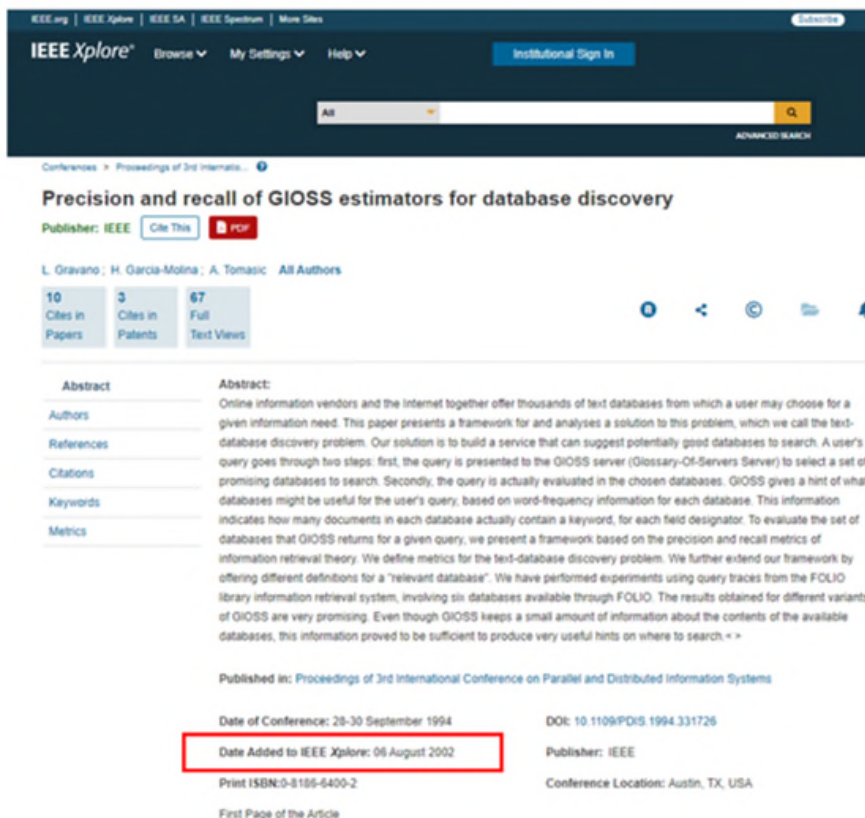
27

8927433.1

31 of 106

IPR2025-00623
UIPATH EXHIBIT 1004

Ex. 1020; Ex. 1025 (Exs. G & H). By 2002, Gravano was also available on the IEEE Xplore website:



Ex. 1021; Ex. 1025 (Exs. A-C). Further, Gravano was cited to the USPTO on March 25, 2005 during the prosecution of RE45752, and a copy of Gravano is part of the publicly available prosecution history thereof. Ex. 1022, at 290-292; Ex. 1024 (Exs. A & B).



PTO/SB/08A (08-03)

Substitute for form 1449A&B/PTO		Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT		Application Number	11/027,585
(use as many sheets as necessary)		Filing Date	December 29, 2004
		First Named Inventor	Gvily, Yaniv
		Art Unit	2171
		Examiner Name	Not Yet Assigned
		Attorney Docket Number	017900-000720US
Sheet	1 of 1		

U.S. PATENT DOCUMENTS+					
Examiner Initials*	Cite No. ¹	Document Number Number Kind Code ² (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear

FOREIGN PATENT DOCUMENTS							
Examiner Initials*	Cite No. ¹	Foreign Patent Document Country Code ³ Number ⁴ Kind Code ⁵ (if known)		Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T ⁶

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials *	Cite No. ¹	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T ²
	AA	GRAVANO, Luis et al.; "Precision and Recall of GIOSS Estimators for Database Discovery"; 1994, <u>Proceedings of the Third International Conference on Parallel and Distributed Information Systems</u> , IEEE Computer Society, pp. 103-106.	
	AB	GRAVANO, Luis et al.; "Precision and Recall of GIOSS Estimators for Database Discovery"; 1994, <u>Stanford University Technical Note No. STAN-CS-TN-94-10</u> , 17 pages.	

Before the priority date of the '679 Patent, Gravano was indexed and catalogued with both an ISBN number and an XP number, available on the IEEE Explorer database, and was not only “locatable” by interested persons, but had actually been found by such persons.

XI. SUMMARY OF GROUNDS FOR UNPATENTABILITY

Ground	Claim(s)	Combinations of References Rendering the Challenged Claims Invalid Under 35 U.S.C. §103
1	1, 2, 8-12, 15-16 and 19-20	Safford, Chen and Laribi
2	3, 13, 17	Safford, Chen and Laribi
3	4, 14, 18	Safford, Chen, Laribi and Siebel

Ground	Claim(s)	Combinations of References Rendering the Challenged Claims Invalid Under 35 U.S.C. §103
4	6-7	Safford, Chen, Laribi, Schultz and Gravano
5	5	Safford, Chen, Laribi, Schultz and Moulinier
6	1, 2, 8-12, 15-16 and 19-20	Safford, Chen and Masters
7	3, 13, 17	Safford, Chen, Masters and Chen
8	4, 14, 18	Safford, Chen, Masters and Siebel
9	6-7	Safford, Chen, Masters, Schultz and Gravano
10	5	Safford, Chen, Masters, Schultz and Moulinier
11	1, 2, 8-12, 15-16 and 19-20	Heath, Randall and Laribi
12	3, 13, 17	Heath, Randall, Laribi and Chen
13	4, 14, 18	Heath, Randall, Laribi and Siebel
14	6-7	Heath, Randall, Laribi, Schultz and Gravano
15	5	Heath, , Randall, Laribi, Schultz and Moulinier

XII. ALL OF THE CHALLENGED CLAIMS ARE OBVIOUS

A. Ground 1 - Claims 1, 2, 8-12, 15-16 and 19-20 Are Obvious Over Safford, Chen and Laribi

1. Claim 1

48. All elements of Claim 1 are met by Safford, Chen and Laribi.

a. Element 1.1 - “A method of analyzing data”

49. To the extent the preamble is limiting, Safford discloses techniques for analyzing social media accounts (which are “data”) associated with a particular entity. Ex. 1006, 3:14-18 (discussing “identifying and analyzing the social signals that are most relevant to the social network performance of a particular entity.”).

b. Element 1.2 - “generating, by an entity, a query based at least in part on a topic of interest”

50. Safford teaches techniques for identifying social media accounts associated with a company or brand, which maps to a “topic of interest.” Ex. 1006, 3:28-36 (“Accounts 108 are stored within analytic system 100 and identify corresponding social network accounts within the social networks 104 ... Accounts 108 also may attempt to identify substantially all of the social network accounts for substantially all of the products marketed by each of the companies”) and 6:3-26. Safford identifies social media accounts associated with a particular company or brand by comparing keywords in messages posted on the social media accounts to “a topic cluster of keywords... derived from the social signals generated by the company account that markets the brand.” *Id.*, 15:39-41. In Safford, “messages” refer to “any communications exchanged via a social network... and any content or information that may be associated with the communication. Ex. 1006, 3:48-54. For example, messages may comprise posts, blogs, Tweets, re-tweets, sentiment indicators, emails, text messages, videos, wall posts, comments, photos, links, or the like, or any combination thereof.” *Id.* The “brand” (or the name of the company that markets the brand or a specific product name) in Safford is “a topic of interest” as recited in the claim.

51. In the ‘679 Patent, a query is described as comprising one or more key words or search terms. Ex. 1001, 7:22-23 (referring to “the key words of the query”), at 7:53-55 (“[a] relevant post refers to data on the data source that match the query and/or key words of the query”), and at 24:1-3 (“according to aspects of the present disclosure the search terms of the query may be expanded”).² Therefore, Safford’s “topic cluster of keywords” for a brand is a “query based at least in part on a topic of interest” for purposes of the ‘679 Patent. A POSITA would have understood such a cluster of key words to include the name of the name brand itself as a keyword, as inclusion of the name of the brand as one of the keywords would be necessary to identify posts mentioning the brand.

52. In the alternative, it would have been obvious based on Chen to include the name of the brand itself as a keyword in order to effectively identify posts discussing the brand. Ex. 1011, 6:32-34 (“for a set of brands, the query should include all the brand names and possible variations of those brand names.”) Chen and Safford both related to methods for monitoring online sentiments associated with a particular product/brand. See, e.g., Ex. 1006, 5:20-23 (“The analytic system may determine that the messages generated by the individual account 108G related

² Safford states that “a topic cluster of keywords may be derived from the social signals generated by the company account that markets the brand.” Ex. 1006, 15:37-41. It would have been obvious for the keywords in Safford to have been derived from the social signals by a person, such as an employee of the company.

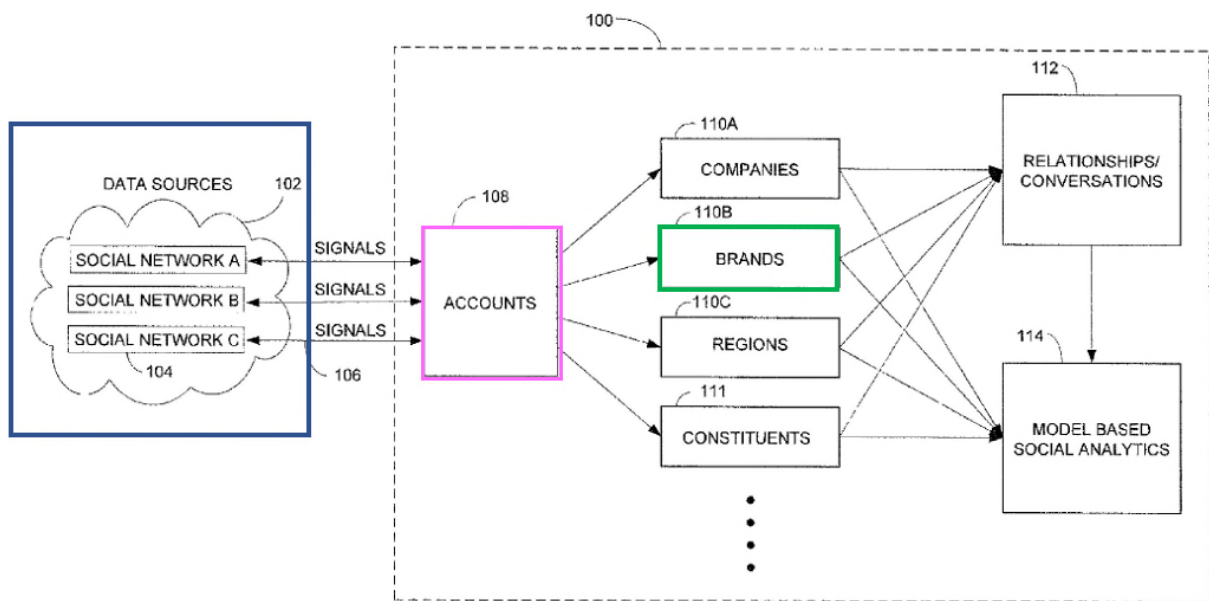
to company account 108E have an overall positive sentiment") and Ex. 1011, 5:11-17 ("[i]n addition to the identity and topic models already discussed, additional models may be used, such as a sentiment model that detects positive and negative tones of text, thereby ascertaining whether a given topic has had a positive or negative effect from the organization's perspective. This allows more precise alerting as well as deeper insights into the brand and reputational issues.") To the extent not already disclosed by Safford, a POSITA would have been motivated to incorporate Chen's technique for using the name of the brand itself as a keyword in Safford's cluster of keywords, as inclusion of the name of the brand as one of the keywords would facilitate the identification of posts mentioning the brand.

53. With respect to the requirement that the query be generated "by an entity," Safford contemplates a platform that analyzes social media information on behalf of a plurality of companies. Ex. 1006, 3:33-36 (the system "may attempt to identify substantially all of the social network accounts for substantially every major company for a variety of different industries."). It would have been obvious for a company to implement Safford's platform internally. The company would have been motivated to implement Safford internally in order to eliminate the need to engage an outside vendor to analyze social media information associated with the company. A company implementing Safford's platform internally maps to "an entity" in the claim.

c. Element 1.3 - “executing the query on a plurality of data sources”

54. Each social media account that is probed by Safford maps to a “data source” in the claim. Ex. 1001, 5:62-64; 15:37-41; 18:17-23. Further, the ‘679 Patent teaches that a blog is an example of a “data source.” Ex. 1001, 4:44-46 and 7:38-42. A social media account (as disclosed in Safford) is analogous to a blog. Both a social media account (*e.g.*, a Twitter account) and a blog will contain a collection of posts (“data,”) and thus constitute a “data source” in the context of the ‘679 Patent. *See* , Ex. 1001, 5:62-64 (referring to “user comments, and/or social media” as data sources.)

55. Safford (alone or combined with Chen) compares keywords that include the name of a brand (“executing a query”) against social media accounts (“a plurality of data sources”) throughout social networks (*e.g.*, Twitter and Facebook) to identify particular social media accounts that refer to the brand (“topic of interest”). Ex. 1006, 3:19-24, 28-36 and 55-67. The process is shown in Safford Figure 2, reproduced below with annotations:



56. In Figure 2 (above), “data sources 102 [blue box] may comprise one or more social networks 104, such as Twitter®, Facebook®, YouTube®, Google+®, or the like.” *Id.*, 3:19-24. By comparing keywords to “posts, blogs, Tweets, re-tweets, sentiment indicators, emails, text messages, videos, wall posts, comments, photos, links, or the like” in the data sources 102 [blue box], Safford identifies particular social media accounts 108 [pink box] that refer to a particular brand 110 [green box]. *Id.*, 3:48-67. Safford identifies social media accounts associated with a particular company or brand by comparing keywords in messages posted on the social media account to “a topic cluster of keywords” associated with the brand. *Id.*, 6:22-28 (“For example, a user of an account may post or send a message that discusses a particular company, brand, etc. The analytic system may compare keywords in the message with keyword topic clusters associated with the company”).

and/or brand") and 15:22-41. Safford's teaching of comparing keywords associated with a company or brand (executing the query) against content posted on social media accounts (plurality of data sources) maps to "executing the query on a plurality of data sources."

d. Element 1.4 - "selecting, by the entity, a data source from the plurality of data sources for monitoring based on a correlation between the data source and the topic of interest, the correlation determined based on results of the executed query"

57. Within the social media accounts that have been associated with a particular company or brand, Safford identifies for monitoring social media accounts (data sources) that correspond to "a source, influencer or advocate for the company" for periodic monitoring, which maps to "selecting, by the entity, a data source from the plurality of data sources for monitoring." More specifically, Safford teaches selecting social media accounts (data sources) that correspond to "a source, influencer or advocate for the company" for periodic monitoring. Ex. 1006, 12:15-20 ("Analysis module 326 may identify the social network account as a source, influencer or advocate for the company and automatically add a new account 108 to the analytic system that identifies the social network account. Collectors 320 then may start periodically collecting signals from the newly identified social network account.") (emphasis added).

58. As explained below, Safford, alone or with Chen, teach that the selection of social media accounts that correspond to a source, influencer or advocate for the company (data sources) is based on a correlation between the data source (the social media account of “an advocate,” “a detractor” or “an influencer” of the company or brand) and the topic of interest (the company or brand), where the correlation is determined based on results of the executed query (whether the social media account has posts with keywords correlated to (or matching) the identity of the company or brand), as claimed.

59. The ‘679 Patent does not expressly define “a correlation between the data source and the topic of interest.” However, the ‘679 Patent teaches that a social media post is determined to be relevant when there is “data on the data source that match the query and/or the keywords of the query.” Ex. 1001, 7:53-55. Thus, selecting a social media account (data source) for monitoring based on whether the social media account includes keywords associated with a company or brand (topic of interest) maps to “selecting ... a data source from the plurality of data sources for monitoring based on a correlation between the data source and the topic of interest.” Like the ‘679 Patent, one criterion used by Safford to select social media accounts for monitoring for a company or brand includes comparing keywords (corresponding to “the topic of interest”) in messages posted on social media accounts to a topic cluster of keywords associated with the company or brand. Thus, Safford performs

“selecting ... a data source from the plurality of data sources for monitoring” based at least in part “on a correlation between the data source and the topic of interest, as claimed.”

60. Safford teaches additional criteria (beyond just a correlation between the data source and the topic of interest)³ are used for determining whether to select a social media account for monitoring, including use of a sentiment analysis to determine whether a given social media account corresponds to “an advocate,” “a detractor” or “an influencer” with respect to a company or brand. Ex. 1006, 16:41-62. Once such a social media account is identified, Safford deploys a “collector” to start “periodically polling “ the account. *Id.*, 6:60-64.

- e. **Element 1.5 - “monitoring, based on a set schedule, the data source for matches to the query to extract data from the data source when at least ... newly added data matches the query”**

Element 1.6 - “extracting data from the data source when at least ... the newly added data matches the query”

³ Claim 1 of the '679 Patent is an open-ended claim, meaning that it allows for unlisted elements to be included within the scope of the claim. As a result, Claim 1 of the '679 Patent encompasses systems like Safford that perform “selecting... a data source from the plurality of data sources for monitoring” based on criteria that include both “a correlation between the data source and the topic of interest” and further criteria such as where the data source corresponds to an “an advocate,” “a detractor” or “an influencer” of the company or brand.

61. As noted, Safford teaches “collectors” that “may be automatically configured to start periodically polling” the selected accounts. Ex. 1006, 12:18-20. “Periodically polling” the selected social media accounts in Safford teaches or renders obvious “monitoring, based on a set schedule, the [selected] data source” in the claim. In addition, Safford’s “collectors” that periodically poll signals from the identified social network accounts function to collect (or “extract”) data resulting from the period polling. Ex. 1006, at 12:15-36. Any data that first appears on one of the identified social network accounts between one polling and the next subsequent polling that extracts such data, constitutes “newly added” data because the data was first added between the two pollings.

62. A POSITA would have understood that Safford (either alone or in combination with Chen) taught “collectors” that periodically poll specific social media accounts representing, e.g., “an advocate,” “a detractor” or “an influencer” of the company or brand, including “extracting” data from messages posted on the social accounts having, among other things, keywords (query) matching the company or brand name. Thus, Safford (either alone or in combination with Chen) teach or render obvious, “monitoring... the data source for matches to the query ... when ... newly added data matches the query” and “extracting data from the data source when ... the newly added data matches the query.”

f. Element 1.7 - “determining an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period”

63. As mentioned above, Safford, alone or in combination with Chen, monitors social media accounts associated with a company or brand and extracts information from them when new messages mentioning the company or brand are posted to the accounts. *See*, discussion of Elements 1.5 and 1.6, *supra*. Safford monitors the rate at which new messages associated with a company or brand are generated. Ex. 1006, 32:37-40 (“[t]he current number of messages generated in ecosystem 492 during a most recent hour, day, week, month, etc. may be identified in current data set 834 and compared with values in historic data set 832 for corresponding time periods.”) However, Safford does not expressly disclose “determining an extraction rate for extracting the data.” This limitation would have been obvious based on Safford and Laribi as explained below.

i. Safford and Laribi Are Both Directed To The Extraction and Subsequent Analysis Of Data In A Cloud Computing Environment

64. Safford discloses two interrelated processes that may be implemented in a “cloud-based computing and storage” environment, namely, the extraction and subsequent analysis of information from social media accounts associated with the company or brand. Ex. 1006, 9:32-34 (discussing “cloud-based environment”); and

13:7-12 (“Collectors 320 may also extract metadata associated with the signals. For example, a Tweet message may contain a mention tag identifying the new phone. The collectors extract and store the mention tag with the Tweet message. The analytic system may use the mention tag to associate the signal with a contextual dimension... “). More generally, Figure 13 of Safford illustrates extraction of data and Figure 18 of Safford describes the subsequent analysis of that data. A POSITA would have understood that Safford’s analysis process was dependent on Safford’s extraction process, because Safford can only analyze data that was previously extracted.

65. Laribi teaches a “performance monitoring service” that, like Safford, monitored performance metrics of “data collection” applications as well as “data analysis” applications in a cloud-based environment. See Ex. 1008, Abstract (referring to “cloud computing environment); Background, 1:21-27; and 11:32-35 (“The performance monitoring service 198 may include monitoring, measurement and/or management software... including data collection, aggregation, analysis, management and reporting”); see also, *Id.*, 12:14-16 and 13:7-24. Laribi taught monitoring such performance metrics in order to provision or deprovision one or more instances of the application. *Id.*, 1:31-46, 1:51-2:2 and 2:7-10. Thus, both Safford and Laribi are directed to data extraction and analysis in a cloud based environment. As explained more fully below in subsection (ii), a POSITA would

have been motivated to combine the teachings of Laribi and Safford in order to monitor the load on Safford's data analysis process and provision or deprovision a number of instances of Safford's data analysis application. For example, a POSITA would have been motivated to provision additional instances of Safford's data analysis application at times when the rate of extracted data (or load) was high, so as to ensure that the extracted data was analyzed in a timely manner.

ii. In View Of Laribi, A POSITA Would Have Been Motivated To Monitor The Load On Safford's Data Analysis Process

66. Among other performance metrics, Laribi teaches monitoring applications based on load. Ex. 1008, 2:19-23 ("the method includes monitoring ... the one or more metrics identifying a load across the plurality of instances of the application.") As noted above, Safford discloses two interrelated processes, namely, the extraction (collection) and subsequent analysis of the extracted information. A POSITA would have understood that the extracted data in Safford corresponded to the load on Safford's subsequent data analysis process.

67. In view of Laribi, a POSITA would have been motivated to monitor the load on Safford's data analysis process (the load corresponding to the rate of extracted data being provided to such data analysis process) for the purpose of provisioning or deprovisioning a number of instances of Safford's data analysis application. A POSITA would have been motivated to provision additional instances

of Safford's data analysis application at times when the rate of extracted data (or load) was high, so as to ensure that the extracted data was analyzed in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of Safford's data analysis application at times when the rate of extracted data (or load) was low in order to conserve computer resources and/or power. Ex. 1008, 58:18-44 and 61:34-43. As explained in Laribi:

In one embodiment of a method of adaptive application provisioning, an appliance ... may monitor ... server or virtual server load, ... An administrator may set a policy identifying a first threshold, such as 80%, above which additional machines may be provisioned or started ... Similarly, the administrator may set a second threshold, such as 20%, below which the number of machines may be reduced or scaled in.... Ex. 1008, 58:18-34 (emphasis added).

* * *

In addition to scaling in or out applications as necessary to service client requests, the above systems and methods may be used to reduce power requirements... Shutting down these machines may allow for conservation of energy when servers are not needed to meet use requirements. *Id.*, 61:34-42 (emphasis added).

68. Safford recognized that the rate at which new messages associated with a company or brand are generated will vary over time. Ex. 1006, 32:37-40 (“[t]he current number of messages generated in ecosystem 492 during a most recent hour, day, week, month, etc. may be identified in current data set 834 and compared with values in historic data set 832 for corresponding time periods.”) Safford also taught that, upon brand sentiment falling below a threshold, a message would be sent to a

manager so that the manager could *launch a social media campaign* to address the sentiment drop. The message would also include a prediction of how much the overall *signal (message) count was predicted to increase*. Ex. 1006, 33:31-45. Safford taught to expect an increase in the number of messages being generated on social media when there were sharp changes in sentiment. *Id.*

69. By 2013, companies understood the need to respond rapidly to issues arising on social media relating to their brands. As explained, *e.g.*, in U.S. Patent Publ. No. 2010/0246797 [Ex. 1027] at [0021], "before a company or entity is even aware of a problem, the viral nature of information spreading on the internet can lead to a ground swell of negative publicity that may or may not be actually based on fact." As a result, it was well-understood that "the time lag between when the information is posted on the internet and what and how a company responds" must be minimized. *Id.* Thus, a company or brand would have been motivated to quickly analyze new messages being generated on social media so that issues relating to the company or brand (*e.g.*, a rapid decline in sentiment) could be identified and addressed rapidly. Based on Safford's insight that variations in the number of messages being generated would be expected to coincide with sharp changes in sentiment, a POSITA would have been motivated to deploy Stafford's platform in a manner that enabled rapid extraction and analysis of new messages generated on

social media at times when the generation rate of such messages was high (*e.g.*, at a time coinciding with a sharp decrease in brand sentiment).

70. Given Safford's insight that the rate at which new messages associated with a company or brand are generated may vary over time, and the motivation of companies to quickly respond to brand issues arising online, a POSITA would have been motivated to deploy Safford's system using a framework such as disclosed in Laribi, that allowed the platform to scale up its capacity to analyze social messages at times when the generation (and extraction) rate of such messages was high, and scale down such capacity when the message generation (and extraction) rate was slow in order to conserve computer resources and/or power, as taught in Laribi. Ex. 1008, 58:18-44 and 61:34-61. Accordingly, it would have been obvious in view of Laribi to determine “an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period” in Safford, so as to monitor the load on Safford’s subsequent data analysis process for the purpose of provisioning or deprovisioning a number of instances of Safford’s data analysis application.

- g. Element 1.8 - “determining a first processing rate for processing the extracted data with a number of parallel processors, the first processing rate indicating an amount of extracted data that is processed over a second time period”**

71. Laribi teaches provisioning/deprovisioning instances of an application based on multiple metrics. Ex. 1008, 1:60-62 (“[t]he method may include monitoring ... each of the one or more metrics for the plurality of instances of the application”). Also, as noted above, Laribi teaches that its “performance monitoring service” monitors performance metrics of both “data collection” and “data analysis” applications. *Id.*, 11:32-35. A POSITA would have understood that monitoring performance of a data analysis application includes monitoring the rate of data processed by the application. Further, in view of Laribi, it would have been obvious to monitor the data processing rate of Safford’s data analysis (or processing) application and adaptively provision instances thereof in order to ensure that the data processing rate of Safford’s data analysis application was sufficient to timely process data being provided thereto by Safford’s data collection (or extraction) process. Accordingly, it would have been obvious in view of Laribi to determine “a first processing rate for processing the extracted data” in Safford where “the first processing rate” indicates “an amount of extracted data that is processed” over a given time.

72. As explained below in connection with Element 1.9, the “second time period” in the claim corresponds to a time when Safford’s data analysis application is performing at a processing rate (corresponding to the “first processing rate” in the claim) that is insufficient to timely analyze the load of extracted data being provided

to Safford's data analysis application by Safford's data collection (extraction) operation. For example, consider the case depicted in Figure 37 of Safford, reproduced below with colored annotations. During Monday and Tuesday, the overall signal count remained below 4.5 (red line) and, in such a case, an operator of the Safford's platform would deploy sufficient resources to analyze a signal count of 4.5 (corresponding to the "first processing rate" in the claim) but would not deploy excess resources for analyzing the signal count as doing so would be a waste of resources and power, as taught in Larbi. Ex. 1008, 61:34-42.

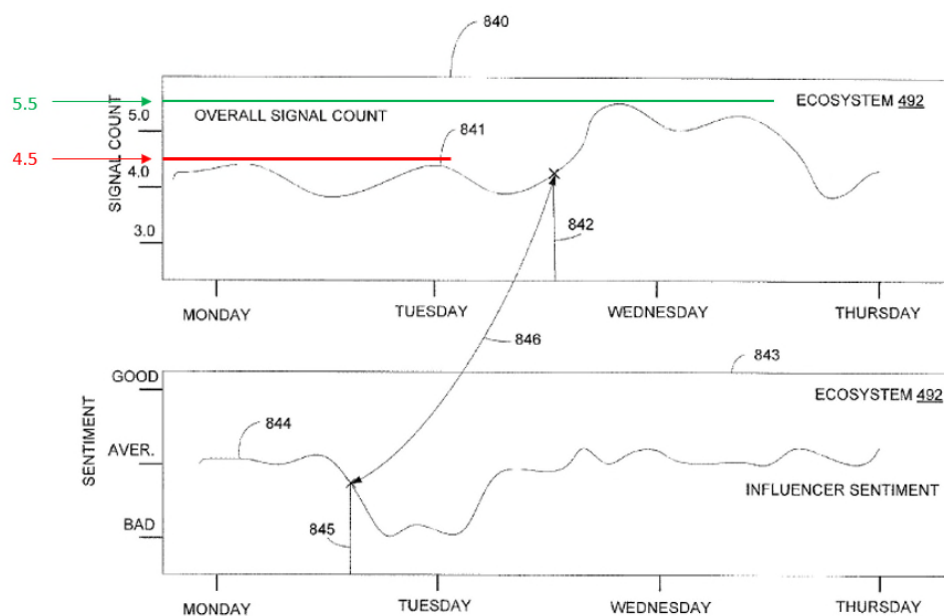


FIG. 37

However, as depicted in the diagram above, there is a rise in the overall signal count to about 5.5 by Wednesday (green line), which coincided with a drop in sentiment the previous day. At that point, the resources previously deployed to analyze an overall signal count of 4.5 (such resources operating at the claimed "first processing

rate") would be insufficient to timely analyze the load of extracted data being provided to Safford's data analysis application by Safford's data collection (extraction) operation given the increased overall signal count from 4.5 to 5.5.

73. Laribi also teaches that, when there are multiple instances of an application, each application can run in parallel on a separate processor. Ex. 1008, 17:6-15 ("the computing device 200 may comprise multiple processors and may provide for simultaneous execution ... In one of these embodiments, the computing device is a shared memory parallel device, with multiple processors") and 39:1-4. Accordingly, in view of Laribi, it would have been obvious to implement Safford's data analysis process by deploying multiple instances of the application on parallel processors which scale in number as the number of instances of the application scale.

h. Element 1.9 "dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate to obtain a second processing rate that is greater than the first processing rate"

74. In view of Laribi, a POSITA would have been motivated to dynamically provision additional instances of Safford's data analysis application on parallel processors at times when Safford's data analysis application was performing at a processing rate (the "first processing rate") that was insufficient to timely analyze the load of extracted data being provided to Safford's data analysis application by Safford's data collection (extraction) operation.

75. Continuing with the example from Figure 37 of Safford (reproduced again below), as explained in connection with element 1.8, there is a rise in the overall signal count to about 5.5 by Wednesday (green line), which coincided with a drop in sentiment the previous day.

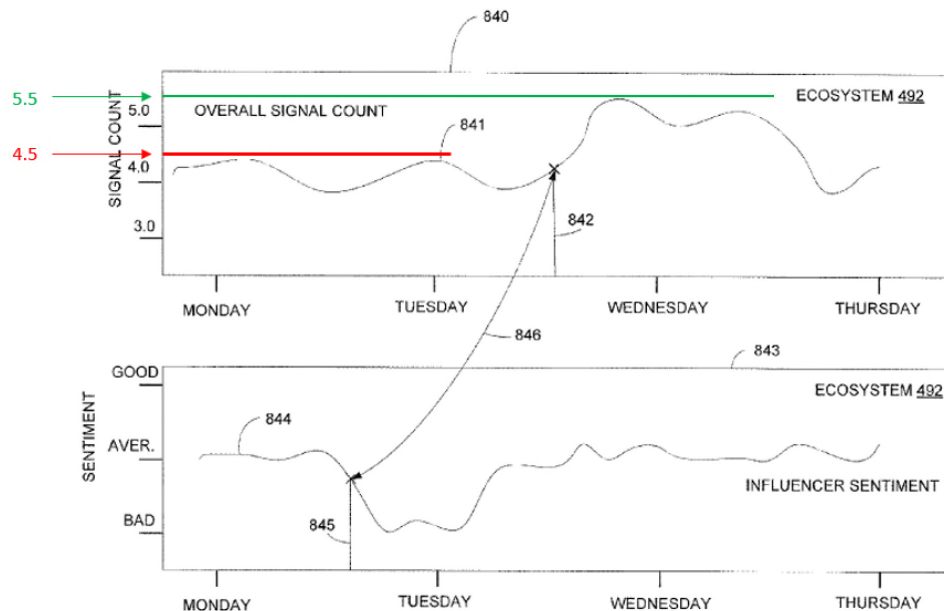


FIG. 37

76. Given that a company or brand would have been motivated to quickly extract and analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly, the company or brand operating Safford's platform would have been motivated to extract the messages at an increased rate to enable rapid analysis of the messages. A POSITA would have looked to Larbi to meet this design goal. More specifically, a POSITA would look to Laribi's scaling techniques for the purpose of provisioning or deprovisioning a number of instances

of Safford's data extraction application. A POSITA would have been motivated to provision additional instances of Safford's data extraction application at times when the rate of data generated on social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford), so as to ensure that the messages being generated on social media were extracted in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of Safford's data extraction application at times when the rate of data generated on social media was low (e.g., an overall signal count of 4.5 in Figure 37 of Safford) in order to conserve computer resources and/or power. Ex. 1008, 58:18-44 and 61:34-43.

77. When the extraction rate in Safford/Laribi combination increased to match an overall signal count of 5.5, the data analysis resources previously deployed in the system to analyze messages extracted at a rate corresponding to a signal count of 4.5 would be insufficient to timely analyze the load of extracted data being provided to Safford's data analysis application. Given that a company or brand would have been motivated to quickly analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly, the company or brand would have been motivated to analyze incoming messages at an increased rate to facilitate a rapid response on social media. Ex. 1006, 33:31-45 (teaching the launching of "a social media campaign ... to address the issues associated with the sentiment drop.")

Again, a POSITA would have looked to Larbi to meet this design goal. More specifically, a POSITA would have looked to Larbi's scaling techniques for the purpose of provisioning or deprovisioning a number of instances of Safford's analysis application. A POSITA would have been motivated to provision additional instances of Safford's data analysis application at times when the rate of data extracted from social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford), so as to ensure that the messages being extracted from social media were analyzed in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of Safford's data analysis application at times when the rate of data being extracted from social media was low (e.g., an overall signal count of 4.5 in Figure 37 of Safford) in order to conserve computer resources and/or power. Ex. 1008, 58:18-44 and 61:34-43.

78. Subsequent to provisioning additional instances of Safford's data analysis application when the rate of data extracted from social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford), (which correlates to "dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate"), the Safford/Larbi system will use a faster data analysis processing rate (or "second processing rate" that is greater than "the first processing rate") to, e.g., timely analyze an overall signal count of 5.5 in Figure 37 of Safford. Ex. 1008, 1:60-62, 11:32-35, 58:18-44 and 61:34-43.

i. Element 1.10 - “analyzing, with the parallel processors, the extracted data to determine at least one of a sentiment ...”

79. Safford teaches analyzing an extracted social media post or account in order to assign a “sentiment score” to the social media post or account. Ex. 1006, 8:43-46, 9:3-5, 16:12-14 and 16:41-45. In the Safford and Laribi combined system discussed in connection with Elements 1.8 and 1.9 above, the analysis is performed on parallel processors. Ex. 1008, 17:6-11.

j. Element 1.11 - “establishing a two-way communication channel, between at least the entity that selected the data source for monitoring and a user device of a user that provided the data to the data source, based on the analysis of the extracted data”

80. Safford’s system classifies certain social media accounts that have been associated with a company into “advocates” or “influencers” based on the analysis of the extracted data. Ex. 1006, 19:58-20:6 (“... constituent classifier 552 may identify some of candidate accounts 490B that generate a certain number of positive messages about the car company and/or the new car sold by the car company as advocates 111A... Candidate accounts 490B having a certain level of social network influence, but not having a particularly strong sentiment regarding the company ecosystem, may be classified as influencers 111D.”) Safford also teaches that “a company can allocate personnel to interact with important advocate accounts” and influencer accounts. *Id.*, 20:14-15 and 20:20-28 (“[t]he coffee company may assign

more personnel to the influencer account 111D in an attempt to convert the influencer into an advocate 111A.”)

81. A POSITA would have understood that “interact[ing] with important advocate accounts” and “influencers” in Safford would have included “establishing a two-way communication channel” between the company or brand and the advocate or influencer (in either case, the “user that provided the data to the data source”) by exchanging messages or posts between the company or brand and the advocate or influencer on the social media platform hosting the advocate or influencer, and/or using email or texts. *See* Ex. 1001, 5:16-17 (referring to a message board and email as examples of communication channels). Such a communication channel would have been required to exchange messages or posts between the company or brand and the advocate or influencer. Social Media platforms (such as the Twitter platform discussed in Safford) had direct messaging functionality that provided an established communication channel through which one user directly communicated with another user. Ex. 1007, 4:66-5:6; Ex. 1029, 14:39-43; Ex. 1030, [0100]; and Ex. 1031, [0025]-[0026]. To the extent it is not explicit in Safford, it would have been obvious for the company or brand to communicate with the advocate or influencer on social media since the advocate or influencer is known to be reachable on social media. The company or brand would motivated to communicate with such an advocate or

influencer, for example, as part of the launching of "a social media campaign ... to address the issues associated with the sentiment drop." Ex. 1006, 33:31-45.

82. In the present mapping, the claimed “entity that selected the data source for monitoring” maps to, *e.g.*, the “coffee company” disclosed in the example (discussed above) and described in Safford. Ex. 1006, 20:20-28. In Safford, the platform performs identification of the advocate/influencer on behalf of the coffee company, and then coffee company then assigns personnel to communicate with the advocate account. As discussed in connection with Element 1.2 above, it would have been obvious, however, for the coffee company to implement Safford’s platform internally, in which case the “two-way communication channel” would be between “the entity (the coffee company) that selected the data source (advocate account) for monitoring and a user device of a user (owner of the advocate/influencer account) that provided the data to the data source.”

- k. **Element 1.12 - “transmitting, from the entity via the two-way communication channel, a first message directed to the user device based on the analysis of the extracted data” and**
Element 1.13 - “receiving, from the user device via the two-way communication channel, a second message in response to the first message directed to the user device”

83. A POSITA would have understood that “interact[ing] with important advocate accounts” or “attempt[ing] to convert the influencer into an advocate” in

Safford would have included transmitting messages to, and receiving responses from, the advocate/influencer via the two-way communication channel (e.g., the social media platform). Ex. 1006, 20:14-15 and 20:20-28. To the extent this is not explicit in Safford, it would have been obvious to interact with advocates and influencers in Safford using a back and forth in which the company or brand reached out to the advocate/influencer (“transmitting ... a first message to the user device”) and the advocate/influencer responded with feedback (“receiving ... a second message in response to the first message”). The company or brand would have been motivated to receive such feedback, for example, as part of the launching of "a social media campaign ... to address the issues associated with the sentiment drop" (ex. 1006, 33:31-45) and/or in order to better service its customer base and in order to build better relationships with advocates/influencers.

84. Claim 1 is therefore obvious over Safford, Chen and Laribi.

2. Claims 11 and 16

85. The additional limitations of Claim 11 (“a memory unit” and “at least one processor ... configured” to perform the recited functions), and Claim 16 (“a non-transitory computer-readable medium having program code recorded thereon” to perform the recited functions), are met by Safford. Ex. 1006, 39:14-25. Thus, Claims 11 and 16 are obvious over Safford, Chen and Laribi for the reasons discussed herein and above with respect to claim 1.

3. Claims 2 and 12 - “visualizing the extracted data”

86. Figure 32 of Safford “depicts an example visualization of a Social Business Graph (SBG) 760 identifying the strength of constituent accounts for a company ... The circles on the graph represent the constituent accounts, such as ... advocate accounts (A), influencer accounts (I), and detractor accounts (D).” Ex. 1006, 29:30-39 (emphasis added). The visualized constituent accounts reflect the “extracted data.” *Id.*, 19:58-60 (discussing “constituent classifier” used to classify candidate accounts). Figures 31 and 33-38 depict other examples where Safford visualizes the extracted data. Thus, Claims 2 and 12 are obvious over Safford, Chen and Laribi.

4. Claim 8 - “the monitoring comprises deploying extractors to the data source, a number of extractors based at least in part on a traffic volume of the data source”

87. As noted above, Laribi discloses a method for automatically scaling a number of instances of an application based on metrics and a threshold. In Laribi, the metric is monitored with respect to the threshold in order to provision or deprovision one or more instances of the application. Ex. 1008, 1:31-46, 1:51-2:2 and 2:7-10. Among other metrics used for scaling up or scaling down instances of an application, Laribi teaches “monitoring ... one or more metrics identifying a load across the plurality of instances of the application.” *Id.*, 2:18-20. In Safford, the load

on the extraction process (performed by “collectors” in Safford) corresponds to the volume of traffic from which data is to be collected. In view of Laribi, it would have been obvious to scale the number of instances of the Safford’s collection application (each of which monitors and extracts data from source(s) of interest) based on the volume of traffic, in order to dynamically adjust the number of instances of the collecting application being deployed “based at least in part on a traffic volume of the data source.” A POSITA would have been motivated to dynamically adjust the number of instances of the collectors being deployed so as to match the extraction capacity of number of instances to the current traffic load, in order to ensure timely extraction of the data. Thus, Claim 8 is obvious over Safford, Chen and Laribi.

5. Claims 9, 15 and 20 - “in which the two-way communication channel comprises at least one of ... social media”

88. As explained in connection with Element 1.11, Safford teaches or renders obvious use of social media as “the two-way communication channel.” Thus, Claims 9, 15 and 20 are obvious over Safford, Chen and Laribi.

6. Claim 10

a. Element 10.1 - “further comprising analyzing, by an entity user corresponding to the entity, the results of the query in view of the topic of interest and a sub-topic of interest”

89. As explained in connection with Element 1.2 above, in Safford the “topic” maps to a “brand or company.” In Safford, the social media account of a specific “advocate” of the brand or company maps to “a sub-topic of interest” in the claim. Ex. 1006, 12:15-20; 16:41-64 Figure 31. Thus, when a representative of the company or brand monitors both the brand as whole on social media as well as the specific social media account of an advocate with respect to the brand, the company or brand representative (“an entity user corresponding to the entity”) is analyzing the results of the query in view of both the topic of interest (the brand as a whole) and a sub-topic of interest (the specific sentiment of an advocate on social media with respect to the brand).

b. Element 10.2 - “determining, by the entity user, the correlation between the data source and the topic of interest based on the analysis of the results of the query”

90. As explained in connection with Element 1.4 above, Safford alone or in combination with Chen teaches selecting data sources (specific social media accounts related to the company or brand) based on a correlation between the data source and the topic of interest (based on whether messages posted on the social accounts have keywords matching the company or brand). Accordingly, claim 10 is obvious over Safford, Chen and Laribi.

7. Claim 19 - “the plurality of data sources are predefined”

91. As explained in connection with Elements 1.3 and 1.4 above, each social media account probed by Safford corresponds to one of a “plurality of data sources,” and based on this probing Safford identifies specific social media accounts for monitoring. Stafford discloses “predefining” the social media accounts to be probed. For example, Safford discloses that “[a]ccounts 108 and signals 106 may be associated with contextual dimensions, such as companies 110A, brands 110B, geographic regions 110C” and “[v]alues of contextual dimensions 110 may be identified *a priori* or ...” Ex. 1006, 3:55-62. In any event, it would have been obvious to predefine a list of social media accounts to be probed before beginning the probing process so that the probing tasks could be efficiently divided in advance amongst the computer resources performing the probing. Thus, Claim 19 is obvious over Safford, Chen and Laribi.

B. Ground 2 - Claims 3, 13 and 17 Are Obvious Over Safford, Chen and Laribi

92. As explained in Sections XII(A)(1) and XII(A)(2), Claims 1, 11 and 16 are obvious over Safford, Chen and Laribi. Claims 3, 13, and 17 depend from Claims 1, 11 and 16, respectively, and contain substantively identical limitations. The analysis below of Claim 3 also applies to Claims 13 and 17.

1. Element 3.1 - “refining search terms of the query based at least in part on the executed query, and in

which the monitoring is performed to monitor for matches to a refined query”

93. The limitations of element 3.1 were obvious over Safford in combination with Chen. Chen disclosed methods for monitoring electronic media and alerting an organization about negative information posted with respect to the organization's brand. Ex. 1011, Title, 1:33-39, and 5:11-16. Chen taught that broad queries concerning a brand may not be “extremely precise” and suggested that a generic brand name such as “‘Dove’® could be refined as ‘Dove’ ® within 10 words of ‘chocolate,” in order to eliminate “irrelevant information.” *Id.*, Title, 1:33-39, 5:11-16 and 6:39-41 (emphasis added). Thus, Chen taught refinement of search terms of the query. Safford and Chen are both directed to methods for monitoring online activity pertaining to a particular brand. A POSITA would have motivated to apply Chen's query refinement techniques to Safford's cluster of keywords for a brand in order to reduce irrelevant information identified by the keywords, and would have had a reasonable expectation of success with respect to the combination. In any event, the concept of refining search queries was notoriously well known by the priority date in 2013. For example, as explained in U.S. Patent Publ. No. 2006/0253427, filed by Google in 2005:

Many users often modify or refine, sometimes repeatedly, their original search query during a given search session. For example, the user may modify an original search query to a more specific search query, a broader search query, and/or a search query using alternative

query terms until desired search results are generated. ... Users typically modify or refine their search queries when the original search query did not yield a good set of search results, e.g., if the search query is too specific or too broad or if the search query used improper terms. For example, the original user search query may yield too many irrelevant results when one or more of the search terms are ambiguous and some of the returned documents are related to a meaning of the ambiguous search term that is different from what the user intended and/or when the user is interested in only one aspect of a given search term among the many aspects of the search term. The original user search query may also yield too many irrelevant results when the user is merely exploring concepts related to a given search term. (Ex. 1028, U.S. Patent Publ. No. 2006/0253427, [0004]).

94. Accordingly, a POSITA would have known to (and been motivated to) modify (or "refine") an original query used in Safford if the query yielded too many irrelevant results or otherwise returned results different from what the user intended.

95. Thus, Claims 3, 13 and 17 are obvious over Safford, Chen and Laribi.

C. Ground 3 - Claims 4, 14 and 18 Obvious Over Safford, Chen and Laribi In View Of Seibel

96. As explained in Sections XII(A)(1) and XII((A)(2), Claims 1, 11 and 16 are obvious over Safford, Chen and Laribi. Claims 4, 14, and 18 depend from Claims 1, 11 and 16, respectively, and contain substantively identical limitations. The analysis below of Claim 4 also applies to Claims 14 and 18.

1. Element 4.1 - “the plurality of data sources include open-source data sources that are publicly available and closed-source data sources that are not publicly accessible”

97. Siebel teaches a text mining system for business intelligence, where the text mining can be performed based on internal corporate documents “such as emails” and “customer support notes.” Ex. 1013, 9:28-29 and 5:33-35. It would have been obvious to apply Safford’s techniques for determining sentiment to a combination of social media channels (which map to “open-source data”) and internal corporate documents “such as emails” and “customer support notes” (which map to “closed-source data sources”) as taught in Siebel. See, e.g., Ex. 1006, 8:43-46, 9:3-5, 16:12-14 and 16:41-45.

98. While many of Safford's embodiments were addressed to social media accounts, Safford also made clear that Safford's techniques were applicable to data sources comprising "any computing system... that generates or aggregates messages that may be exchanged or reviewed by different users." Ex. 1006, 3:25-27. The internal corporate databases used for text mining of “emails” and “customer support notes” in Siebel (ex. 1013, 9:28-29 and 5:33-35) are computer systems that generate or aggregate "messages that may be exchanged or reviewed by different users" as disclosed in Safford. Ex. 1006, 3:25-27.

99. A POSITA would have been motivated to integrate the internal data sources disclosed in Siebel with Safford’s social media data sources for purposes of both the initial query and ongoing monitoring steps recited in the independent claims. Among other reasons, a POSITA would have been motivated to integrate

such data sources in order to expand the scope of data being analyzed and more accurately quantify sentiment based on a combination of social media posts from a particular user (e.g., an advocate) and internal corporate documents “such as emails” and “customer support notes” representing private communications between the company and such user. In this regard, Siebel taught the derivation of business intelligence from multiple sources including “web-based discussion forums” and a corporation’s “internal documents.” Ex. 1013, 5:8, 5:33 and 9:18-30. The same keywords (query) executed against the social media posts in Safford would be applied to the internal corporate data sources identified in Siebel. With respect to ongoing monitoring, it would have been obvious that when a user on social media is identified as an advocate and selected for monitoring in Safford, future social media posts as well as future emails from the user or future support notes (as disclosed in Siebel) associated with the same user would be selected in the Safford/Chen/Laribi/Seibel system for ongoing monitoring. Thus, Claims 4, 14 and 18 are obvious over Safford, Chen, Laribi and Seibel.

D. Ground 4 - Claims 6-7 Are Obvious Over Safford, Chen, Laribi, Schultz And Gravano

1. Claim 6

100. As explained in Section XII(A)(1), Claim 1 is obvious over Safford, Chen and Laribi. Claim 6 depends from Claim 1. The additional limitations of Claim 6 were obvious over Safford and Schultz.

a. Element 6.1 - “expanding search terms of the query”

101. Schultz taught expansion of query terms using a semantic network. Ex. 1012, Abstract. Safford and Schultz are both directed to search query technology. It would have been obvious to apply Schultz’s query expansion techniques to Safford’s “topic cluster of keywords” in order to generate a more robust cluster of keywords.

b. Element 6.2 - “the data source is selected when the correlation is greater than or equal to an accuracy threshold”

102. Safford taught use of a “first threshold” that measured the correctness (or accuracy) of whether or not to associate a particular social media account with a company or brand:

Operation 442 may determine if the number of signals generated by the Twitter® account is above a first threshold. For example, the analysis module may identify the number of signals generated by the Twitter® account during the conversation ... The Twitter® account might not be added to the analytic system if the number of signals is below the first threshold. Ex. 1006, 16:27-34 (emphasis added).

The first threshold in Safford maps to an accuracy threshold, in that the accuracy of whether the social media account is associated with a given company or brand is higher in cases where the first threshold is exceeded. Thus, Safford selects a data source (e.g, a particular social media account) for monitoring when the number of signals matching the keywords (the correlation) exceeds a first threshold (accuracy

threshold). Thus, Claim 6 would have been obvious over Safford, Chen, Laribi and Schultz.

103. Alternatively, element 6.2 would have been obvious over Safford in combination with Gravano (Ex. 1015). Gravano disclosed a metric referred to as " $P_{right}(q, db)$ " which measured, for a given database (db) and query (q), the percentage of retrieved documents from the database that were actually relevant to the query. Ex. 1017, 3. For example, if the query (q) returned 100 documents from a given database (db), and 99 of the documents were relevant (e.g., there was one false positive), the value of $P_{right}(q, db)$ would be 0.99. The P_{right} value is an accuracy metric that measures the precision with which relevant documents are identified.

104. Gravano described users seeking to retrieve information from database collections as being either "recall-oriented" or "precision-oriented." Recall-oriented users want an exhaustive answer to their query, and are willing to accept that a higher proportion of retrieved documents will be useless in order to assure that a higher percentage of all of the relevant documents (throughout all the database collections) are retrieved. Ex. 1017, 2-3. On the other hand, precision-oriented users are willing to accept that a lower proportion of relevant documents (throughout all of the database collections) will be retrieved in order to assure that a higher percentage of documents that are retrieved are useful. Ex. 1017, 2-3. Each user can be thought

of as being on a continuum that reflects the user's preference to trade-off recall for precision. Ex. 1017, 2-3.

105. In the context of collection selection, Gravano defined two functions:

$$P_{Right}(q, DB) = \begin{cases} \frac{|Chosen \cap Right|}{|Chosen|} & \text{if } |Chosen| > 0 \\ 1 & \text{otherwise} \end{cases}$$
$$R_{Right}(q, DB) = \begin{cases} \frac{|Chosen \cap Right|}{|Right|} & \text{if } |Right| > 0 \\ 1 & \text{otherwise} \end{cases}$$

where, for a given query (q) and database (db), $Right$ is the set of items in the database (db) that are relevant to the query (q) and $Chosen$ is the set of items actually retrieved. " P is the fraction of selected databases that are $Right$ ones and R is the fraction of the $Right$ databases that are selected." Ex. 1017, 3. As Gravano explained, "[p]recision oriented users will be interested in high values of P while recall oriented users will be interested in high values of R ." *Id.* (emphasis added).

106. Safford and Gravano are both directed to search techniques, and it would have been obvious for a company implementing Safford's platform to apply Gravano's collection selection metrics to identify social network accounts for monitoring as taught in Safford. See, e.g., Ex. 1006, 12:15-20. For example, a precision-oriented company monitoring its brand using Safford's platform would have be motivated to use a high value of P in connection with selection of data sources in order to reflect the company's preference for precision-oriented searching, as taught in Gravano. Each social network account in Safford corresponds to a

database (*db*) in Gravano, and a precision-oriented company wishing to monitor its brand would have been motivated to identify for monitoring social media accounts having a high value of P in order to lessen the proportion of useless items from the social media accounts being monitored, as taught in Gravano, and would have had a reasonable expectation of success in implementing the combination. The minimum acceptable value of P_{right} used by the precision-oriented company to select data sources corresponds to an "accuracy threshold" (because it reflects a level of search precision) as well as a correlation between the database and the query that is "determined based on the results of the query" (because, the higher the P_{right} value, the higher the correlation between the database and the query.) A precision-oriented user operating Safford's platform using these teachings from Gravano will select data sources for monitoring "when the correlation [$P_{right}(q, db)$] is greater than or equal to an accuracy threshold [the minimum acceptable value of P selected by a precision-oriented user]," as recited in claim 6. Thus, in the alternative, Claim 6 would have been obvious over Safford, Chen, Laribi, Schultz and Gravano.

2. Claim 7 - "expanding the query comprises expanding the search terms to include at least misspellings, synonyms, sub-topics, antonyms, or a combination thereof"

107. As explained in Sections XII(D)(1), Claim 6 is obvious over Safford, Chen, Laribi, Schultz and Gravano. Claim 7 depends from Claim 6. The additional limitation of Claim 7 is obvious over Schultz, which teaches use of a semantic

network which includes “synonym information” for expanding query terms. Ex. 1012, 23:44-48. Thus, Claim 7 would have been obvious over Safford, Chen, Laribi, Schultz and Gravano.

E. Ground 5 - Claim 5 Is Obvious Over Safford, Chen and Laribi, In View Of Schultz and Moulinier

108. As explained in Section XII(A)(1), Claim 1 is obvious over Safford, Chen and Laribi. Claim 5 depends from Claim 1. The additional limitations of Claim 5 are obvious over Safford, Schultz and Moulinier.

1. Element 5.1 - “the query is a natural language query”

109. Schultz teaches that the “search query is preferably in a natural language format.” Ex. 1012, 4:8-9. It would have been obvious to substitute a natural language query for Safford’s keyword clusters, as a natural language query is simpler to formulate.

2. Element 5.2 - “translating the query to a system understood language”

110. Moulinier teaches translation of a natural language query from one language to another. Ex. 1014, 4:10-14 and 6:3-57; *see also* Ex. 1001 at 9:62-64 (“The language translator 408 may translate the query elements and/or the results from English to another language.”) Safford, Schultz and Moulinier are all directed to search query technology. It would have been obvious to apply Moulinier’s

language translation techniques to Schultz’s natural language query.⁴ In a case where Safford’s social media accounts were in different languages, a POSITA would have been motivated to use Moulinier’s language translation techniques with Schultz/Safford, so as to permit the user to search social media accounts in different languages based on a natural language query in the user’s native language.

111. Thus, Claim 5 is obvious over Safford, Chen and Laribi, in view of Schultz and Moulinier.

F. Ground 6 - Claims 1, 2, 8-12, 15-16 and 19-20 are Obvious Over Safford, Chen and Masters

1. Claim 1

112. As explained in Section XII(A)(1), Elements 1.1-1.6 and 1.11-1.13 of Claim 1 are met by Safford alone or in combination with Chen. In Ground 5, the remaining elements of Claim 1 (i.e., Elements 1.7-1.10) are shown as being met by Safford in combination with Masters.

a. Element 1.7 - “determining an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period”

i. Safford and Masters Are Both Directed To Computer Systems That Implement Interrelated And Dependent Applications

⁴ See, Ex. 1001, 9:62-64 (discussing translation of query “from English to another language, and vice versa.”)

113. Both Safford and Masters are directed to computer systems that implement interrelated and dependent applications. As explained in Section XII(A)(1)(f), Safford discloses two interrelated processes, namely, the extraction and subsequent analysis of information from social media accounts. Masters also taught a computer system for managing resources associated with dependent software applications. Ex. 1010, 41:3-9 (“In order to effectively manage a pool of computing resources, the Resource Manager FG42 requires some means or mechanism of determining ... the dependencies of these software components on both hardware and software resources”) and 40:43-46. Thus, both references are directed to a common field of endeavor. A POSITA would have found the teachings of Masters relating to allocating computer resources in the context of dependent applications to be pertinent to dependent applications like Safford’s data extraction and subsequent analysis processes. The motivation to combine the references is discussed next.

ii. A POSITA Would Have Been Motivated To Monitor The Rate Of Extracted Data Being Provided To Safford’s Data Analysis Process In Order To Scale Up Or Scale Down Instances Of Safford’s Data Analysis Application

114. As noted above, Safford recognized that the rate at which new messages associated with a company or brand are generated will vary over time. Ex. 1006, 32:37-40 (“[t]he current number of messages generated in ecosystem 492 during a”
70

most recent hour, day, week, month, etc. may be identified in current data set 834 and compared with values in historic data set 832 for corresponding time periods.”) Safford also taught that, upon brand sentiment falling below a threshold, a message would be sent to a manager so that the manager could *launch a social media campaign* to address the sentiment drop. The message would also include a prediction of how much the overall *signal (message) count was predicted to increase*:

Based on relationship 846, the social analytic system may monitor data set 844 and send messages to the company operating ecosystem 492 whenever the influencer sentiment drops below a threshold amount. The message may identify the sharp decrease in influencer sentiment at time 845 and may also include a prediction indicating when and how much the overall signal count is predicted to increase at time 842.

The account manager for the company operating ecosystem 492 may review messages sent from influencers around time 845 to determine why the sentiment value decreased. A social media campaign can then be launched to address the issues associated with the sentiment drop. For example, the company operating ecosystem 492 may send messages from their social website account addressing a product defect being discussed in the influencer messages. Ex. 1006, 33:31-45 (emphasis added).

Thus, Safford taught to expect an increase in the number of messages being generated on social media when there were sharp changes in sentiment.

115. Furthermore, as explained above, by 2013, corporations understood the need to respond rapidly to issues arising on social media relating to their brands. See, e.g., Ex. 1027, at [0021] (explaining that "before a company or entity is even

aware of a problem, the viral nature of information spreading on the internet can lead to a ground swell of negative publicity that may or may not be actually based on fact" and teaching that "the time lag between when the information is posted on the internet and what and how a company responds" must be minimized.) Thus, a company or brand would have been motivated to quickly analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly. Based on Safford's insight that variations in the number of messages being generated would be expected to coincide with sharp changes in sentiment, a POSITA would have been motivated to deploy Stafford's platform in a manner that enabled rapid extraction and analysis of new messages generated on social media at times when the generation rate of such messages was high (e.g., at a time coinciding with a sharp decline in brand sentiment).

116. Masters disclosed that its QoS managers were responsible for "recommending that applications be either scaled up, scaled down, or moved in order to maintain acceptable performance" and its Resource Manager was responsible for "determining if and where to place new copies of scalable applications or which scalable applications should be shutdown when the QoS Managers indicate that scale-up or scale-down actions should be taken based on measured application performance." Ex. 1010, 18:45-47, 53-58 and 21:40-49

(emphasis added). Given Safford's insight that the rate at which new messages associated with a company or brand are generated may vary over time, and the motivation of companies to quickly respond to brand issues arising online, a POSITA would have been motivated to deploy Safford's system using a framework such as disclosed in Masters, that allowed the platform to scale up its capacity to analyze social messages at times when the generation (and extraction) rate of such messages was high (thus maintaining an acceptable level of performance in which incoming messages were timely analyzed), and scale down such capacity when the message generation (and extraction) rate was slow in order to, e.g., conserve computer resources and/or power.

117. In Masters, “[i]f an application is not meeting its performance requirements... FG44N will either request that the Resource Manager FG42 scale up a new copy of the application or ...”. *Id.*, 21:40-49 (emphasis added). In view of Masters, it would have been obvious to assign a QoS requirement to Safford’s data analysis process requiring Safford’s data analysis process to process (analyze) data at a rate that was at least as fast as Safford’s collectors (“extractors”) supplied extracted data to Safford’s data analysis operation. In order to ensure that this QoS requirement was met, a POSITA would have been motivated to monitor the rate of

extracted data being provided to Safford's data analysis process⁵ for the purpose of scaling-up or scaling down a number of instances of Safford's data analysis application. A POSITA would have been motivated to scale-up additional instances of Safford's data analysis application when the rate of extracted data was higher than the processing rate of the data analysis operation, so as to ensure that the extracted data was analyzed in a timely manner (e.g., to ensure that Safford's data analysis operation was processing data at least as fast as the rate that Stafford's data collectors were extracting data.) Similarly, a POSITA would have been motivated to scale-down instances of Safford's data analysis application when the rate of extracted data was low in order to conserve computer resources and/or power. Accordingly, it would have been obvious in view of Masters to determine "an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period" in Safford, so as to ensure that the extracted data was analyzed in a timely manner for the purpose of provisioning or deprovisioning a number of instances of Safford's data analysis application.

b. Element 1.8 - "determining a first processing rate for processing the extracted data with a number of parallel processors, the first processing rate indicating an amount of

⁵ Such monitoring of the rate of extracted data being provided to Safford's data analysis process maps to "determining an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period."

extracted data that is processed over a second time period”

118. In view of Masters, it would have been obvious to monitor the data processing rate of Safford’s data analysis (or processing) application and adaptively provision instances thereof in order to ensure that the data processing rate of Safford’s data analysis application was sufficient to timely process data being provided thereto by Safford’s data collection (or extraction) process. Accordingly, it would have been obvious in view of Masters to determine “a first processing rate for processing the extracted data” in Safford where “the first processing rate” indicates “an amount of extracted data that is processed” over a given time.

119. As explained more fully below in connection with Element 1.9 below (Section XII(F)(1)(c)), the “second time period” in the claim corresponds to a time when Safford’s data analysis application is performing at a processing rate (corresponding to the “first processing rate” in the claim) that is insufficient to timely analyze the extracted data being provided to Safford’s data analysis application by Safford’s data collection (extraction) operation (e.g., when the extraction rate is greater than the processing rate of the data analysis operation.) For example, consider the case depicted in Figure 37 of Safford, reproduced below with colored annotations. During Monday and Tuesday, the overall signal count remained below 4.5 (red line) and, in such a case, an operator of the Safford’s platform would deploy sufficient resources to analyze a signal count of 4.5 (corresponding to the “first

processing rate” in the claim) but would not deploy excess resources for analyzing the signal count as doing so would be a waste of resources and power.

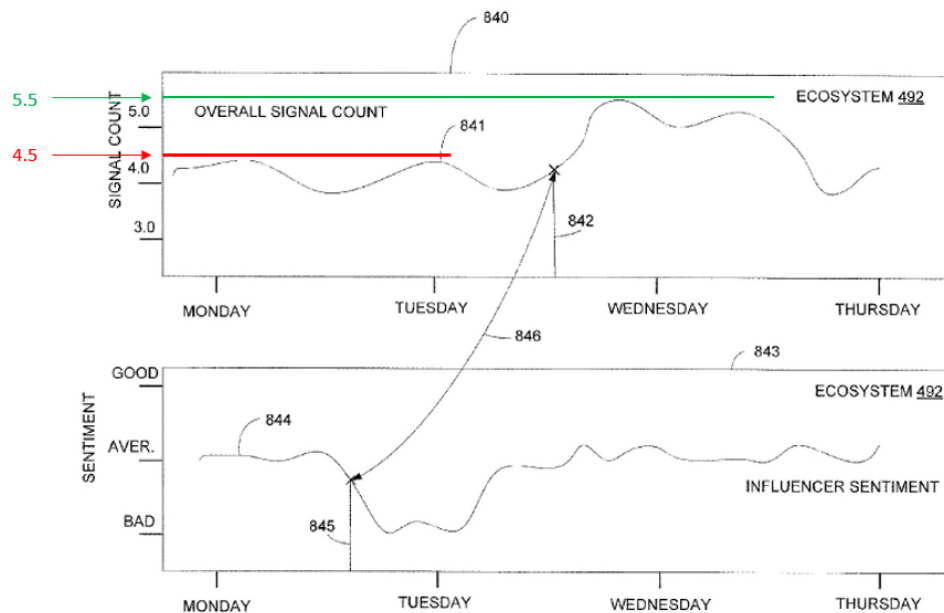


FIG. 37

However, as depicted in the diagram above, there is a rise in the overall signal count to about 5.5 by Wednesday (green line), which coincided with a drop in sentiment the previous day. At that point, the resources previously deployed to analyze an overall signal count of 4.5 (such resources operating at the claimed "first processing rate") would be insufficient to timely analyze the load of extracted data being provided to Safford’s data analysis application by Safford’s data collection (extraction) operation given the increased overall signal count from 4.5 to 5.5.

120. Masters disclosed that its QoS managers were responsible for “recommending that applications be either scaled up, scaled down, or moved in order to maintain acceptable performance” and its Resource Manager was

responsible for “determining if and where to place new copies of scalable applications or which scalable applications should be shutdown when the QoS Managers indicate that scale-up or scale-down actions should be taken based on measured application performance.” Ex. 1010, 18:45-47 and 53-58. Masters describes that its Resource Manager determines "where to place new copies of scalable application," which means that Masters contemplates running different instances of a scalable application on different devices. Ex. 1010, 18:53-61.

121. As of the priority date of the '679 Patent, performing operations on parallel processors was a well-known technique for increasing throughput. See, e.g., Ex. 1008, 17:6-15. Therefore, it would have been obvious to run each of the multiple copies of the scalable application in Masters on a separate physical processor in order to increase throughput when new copies of the application are deployed. See, Ex. 1010, 1:51 and 2:12-13.

c. Element 1.9 “dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate to obtain a second processing rate that is greater than the first processing rate”

122. In view of Masters, a POSITA would have been motivated to dynamically scale-up/scale-down instances of Safford’s data analysis application on parallel processors when Safford’s data analysis application was performing at a processing rate (the “first processing rate”) that was insufficient to keep up with rate

of extracted data input into Safford's data analysis operation by Safford's data collection (extraction) operation.

123. Continuing with the example from Figure 37 of Safford (reproduced again below), as explained in connection with element 1.8 immediately above, there is a rise in the overall signal count to about 5.5 by Wednesday (green line), which coincided with a drop in sentiment the previous day.

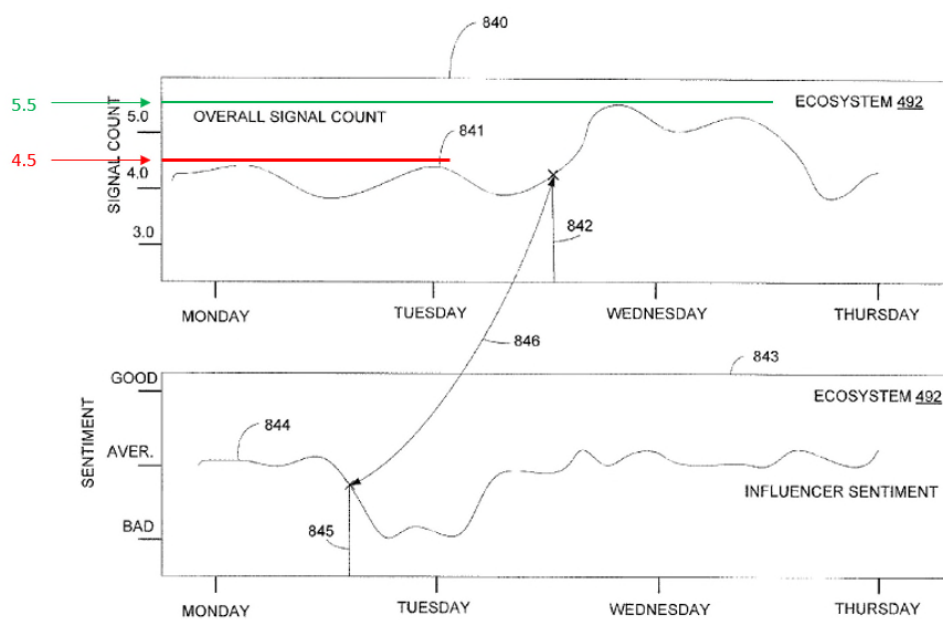


FIG. 37

124. Given that a company or brand would have been motivated to quickly extract and analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly, the company or brand would have been motivated to extract the messages at an increased rate to enable rapid analysis of the messages. A POSITA would have looked to Masters to meet this design goal. More

specifically, a POSITA would look to Master's scaling techniques for the purpose of provisioning or deprovisioning a number of instances of Safford's data extraction application. A POSITA would have been motivated to provision additional instances of Safford's data extraction application at times when the rate of data generated on social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford), so as to ensure that the messages being generated on social media were extracted in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of Safford's data extraction application at times when the rate of data generated on social media was low (e.g., an overall signal count of 4.5 in Figure 37 of Safford) in order to conserve computer resources and/or power.

125. When the extraction rate in the Safford/Masters combination increased to match an overall signal count of 5.5, the data analysis resources previously deployed in the system to analyze messages extracted at a rate corresponding to a signal count of 4.5 would be insufficient to timely analyze the load of extracted data being provided to Safford's data analysis application. Given that a company or brand would have been motivated to quickly analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly, the company or brand would have been motivated to analyze incoming messages at an increased rate to facilitate a rapid response on social media. Ex. 1006, 33:31-45 (teaching the launching of "a

social media campaign ... to address the issues associated with the sentiment drop.") Again, a POSITA would have looked to Masters to meet this design goal. More specifically, a POSITA would have looked to Master's scaling techniques for the purpose of provisioning or deprovisioning a number of instances of Safford's analysis application. A POSITA would have been motivated to provision additional instances of Safford's data analysis application at times when the rate of data extracted from social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford) , so as to ensure that the messages being extracted from social media were analyzed in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of Safford's data analysis application at times when the rate of data being extracted from social media was low (e.g., an overall signal count of 4.5 in Figure 37 of Safford) in order to conserve computer resources and/or power.

126. Subsequent to dynamically scaling up the number of instances of the data processing application when the rate of data extracted from social media was high (e.g., an overall signal count of 5.5 in Figure 37 of Safford) (which maps to "dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate"), the Safford/Master system will achieve a faster data analysis processing rate (or "second processing rate") that is greater than "the first processing rate."

i. Element 1.10 - “analyzing, with the parallel processors, the extracted data to determine at least one of a sentiment, an index, a pattern, or a combination thereof”

127. Safford teaches analyzing an extracted social media post or account in order to assign a “sentiment score” to the social media post or account. Ex. 1006, 8:43-46, 9:3-5, 16:12-14 and 16:41-45. In the Safford/Masters system discussed in Sections XII(F)(1)(b) and XII(F)(1)(c) above, the analysis is performed on parallel processors.

128. Claim 1 is therefore obvious over Safford, Chen and Masters.

2. Claims 2, 11, 12 and 16

129. As explained in Sections XII(A)(2) and XII(A)(3), the additional limitations of Claims 2, 11, 12 and 16 are met by Safford. Ex. 1006, 39:14-25, Figure 32 and 29:30-39. Thus, Claims 2, 11, 12 and 16 are obvious over Safford, Chen and Masters.

3. Claim 8

130. As noted in Sections XII(F)(1)(a)-(c) above, Masters discloses a method for automatically scaling a number of instances of an application based on a QoS threshold. In Safford, the load on the extraction process (performed by “collectors” in Safford) corresponds to the volume of traffic from which data is to be collected. In view of Masters, it would have been obvious to scale the instances

of the Safford's collection application based on traffic volume, in order to adjust the number of instances of the collecting application being deployed "based at least in part on a traffic volume of the data source." A POSITA would have been motivated to dynamically adjust the instances of the collectors (each of which monitors and extracts data from source(s) of interest) being deployed so as to match the extraction capacity of number of instances to the current traffic load, in order to ensure timely extraction of the data. Thus, Claim 8 is obvious over Safford, Chen and Masters.

4. Claims 9, 10, 15, 19 and 20

131. As explained in Section XII(A)(5) above, Safford teaches or renders obvious use of social media as "the two-way communication channel" in Claims 9, 15 and 20.

132. As explained in Section XII(A)(6) above, the additional limitations of Claim 10 are met by, or obvious in view of, Safford.

133. As explained in Section XII(A)(7) above, the additional limitations of Claim 19 are met by, or obvious in view of, Safford.

134. Thus, Claims 9, 10, 15, 19 and 20 are obvious over Safford, Chen and Masters.

G. Ground 7 - Claims 3, 13 and 17 Are Obvious Over Safford, Chen and Masters

135. As explained in Sections XII(F)(1) and XII((F)(2), Claims 1, 11 and 16 are obvious over Safford, Chen and Masters. Claims 3, 13, and 17 depend from Claims 1, 11 and 16, respectively. As explained in Sections XII(B), the additional limitations of Claims 3, 13 and 17 are met by Chen, and it would have been obvious to apply Chen's query refinement techniques to Safford's cluster of keywords for a brand in order to reduce irrelevant information identified by the keywords. Thus, Claims 3, 13 and 17 are obvious over Safford, Chen and Masters.

H. Ground 8 - Claims 4, 14 and 18 Obvious Over Safford, Chen and Masters In View Of Seibel

136. As explained in Sections XII(F)(1) and XII((F)(2), Claims 1, 11 and 16 are obvious over Safford, Chen and Masters. Claims 4, 14, and 18 depend from Claims 1, 11 and 16, respectively. As explained in Sections XII(C), the additional limitations of Claims 4, 14 and 18 are met by Seibel, and it would have been obvious to apply Safford's techniques for determining sentiment to a combination of social media channels ("open-source data") and internal corporate documents "such as emails" and "customer support notes" ("closed-source data sources) as taught in Siebel. Thus, Claims 4, 14 and 18 are obvious over Safford, Chen, Masters and Seibel.

I. Ground 9 - Claims 6-7 Are Obvious Over Safford, Chen, Masters, Schultz and Gravano

1. Claim 6

137. As explained in Section XII(F)(1), Claim 1 is obvious over Safford, Chen and Masters. Claim 6 depends from Claim 1. As explained in Section XII(D)(1), the additional limitations of Claim 6 were obvious over Safford, Schultz and Gravano, and it would have been obvious to apply Schultz's query expansion techniques to Safford's "topic cluster of keywords" in order to generate a more robust cluster of keywords. Thus, Claim 6 would have been obvious over Safford, Chen, Masters, Schultz and Gravano.

2. Claim 7

138. As explained in Section XII(I)(1), Claim 6 is obvious over Safford, Chen, Masters, Schultz and Gravano. Claim 7 depends from Claim 6. As explained in Section XII(D)(2), the additional limitation of Claim 7 was obvious over Schultz. Thus, Claim 7 would have been obvious over Safford, Chen, Masters, Schultz and Gravano.

J. Ground 10 - Claim 5 Is Obvious Over Safford, Chen and Masters, In View Of Schultz and Moulinier

139. As explained in Section XII(F)(1), Claim 1 is obvious over Safford, Chen and Masters. Claim 5 depends from Claim 1. As explained in Section XII(E), the additional limitations of Claim 5 were obvious over Safford, Schultz and

Moulinier. Thus, Claim 5 is obvious over Safford, Chen and Masters, in view of Schultz and Moulinier.

K. Ground 11 - Claims 1, 2, 8-12, 15-16 and 19-20 Are Obvious Over Heath, Randall and Laribi

1. Claim 1

140. All elements of Claim 1 are met by Heath, Randall and Laribi.

a. Element 1.1 - “A method of analyzing data”

141. To the extent the preamble is limiting, Heath discloses “electronically collecting and analyzing ... online social media.” Ex. 1007, 1:30-44.

b. Element 1.2 - “generating, by an entity, a query based at least in part on a topic of interest”

142. Heath teaches the identification of keywords expressing consumer or brand sentiment about terms associated with consumer products or services. Ex. 1007, at 5:31-43.⁶ Identifying keywords expressing consumer or brand sentiment about terms associated with consumer products or services in Heath maps to “generating ... a query based at least on a topic of interest.” With respect to the requirement that the query be generated “by an entity,” Heath contemplates implementation of its platform in “a system internal to and/or controlled by a

⁶ Heath describes the use of “data sets comprising terms or phases of interest” (keywords) but provides no disclosure of an automated method for deriving such keywords. Accordingly, a POSITA would understand or, alternatively it would have been obvious, for the keywords in Heath to have been derived by a human.

particular organization.” *Id.*, 49:4-5. A company implementing Heath’s platform internally maps to “an entity.”

c. Element 1.3 - “executing the query on a plurality of data sources”

143. Heath discloses a “forum analyzer” which “scours the Internet searching for online consumer feedback for products or services via online activity ... and social media content (threads) relevant to a ... product/service ... the forum analyzer 127 accomplishes this using automated tools for identifying industry-specific social media data sources from which to collect information and provide to the users of the CF-SMA platform. This includes a forum analysis to locate or discover which forums and/or sub-forums are relevant to a specific user’s industry ... To accomplish this, search results from publicly available online search engines are processed to determine relevant websites based on the relevance score of each site for the keywords of interest.” Ex. 1007, 54:12-40. The social media data sources, such as forums, sub-forums and websites “scoured” by Heath’s “forum analyzer” map to “a plurality of data sources.” When Heath’s forum analyzer scours such data sources based on “the keywords of interest,” Heath’s forum analyzer is “executing the query” (i.e., the keywords of interest) on “a plurality of data sources.”

d. Element 1.4 - “selecting, by the entity, a data source from the plurality of data sources for monitoring based on a correlation between the data source and the topic of interest, the

correlation determined based on results of the executed query”

Element 1.5 - “monitoring, based on a set schedule, the data source for matches to the query to extract data from the data source when at least ... newly added data matches the query”

144. The "forums and/or sub-forums" that are located as being "relevant to a specific user's industry" by Heath's automated tools are data sources that are selected “based on the relevance score of each site for the keywords of interest.” Ex. 1007, 54:12-40. The “relevance score” in Heath is a measure of the “correlation” between a given data source (e.g., a website) and “the topic of interest.” Heath’s “keywords of interest” correspond to “the topic of interest.”

145. Heath discloses “[m]onitoring and analyzing” each "new information source" and outputting alerts “when consumer or brand sentiment trends change ... allow[ing] for constant, real-time monitoring of emerging trends and consumer or brand sentiment.” Ex. 1007, 30:1-19; 52:8-25, 59:67-60:5; 64:8-10 (disclosing trend line graph showing how attitudes or opinions of brand have changed over time.) It would have been obvious to configure Heath's platform to monitor the "forums and/or sub-forums" that were located as "relevant to a specific user's industry" according to a set schedule. *Id.*, 53:19-21 (referring to "any frequency such as daily, hourly, weekly, and minute-by-minute.”) The monitoring rate is a matter of design choice.

146. To the extent that a monitoring rate was not made obvious by Heath, U.S. Patent No. 7,725,452 (Randall) disclosed crawlers or robots for detecting changes in online content, and the scheduling of crawling "every few minutes, or N times per hour." Ex. 1023, 2:66-3:1, 5:50-59, 6:1. Randall and Heath are both directed to monitoring changes to online content, and a POSITA would have been motivated to deploy crawlers or bots with Heath at regular intervals in order to establish a monitoring rate as taught by Randall. As noted above, by 2013, corporations understood the need to respond rapidly to issues arising on social media relating to their brands. See, e.g., Ex. 1027, at [0021] (explaining that "before a company or entity is even aware of a problem, the viral nature of information spreading on the internet can lead to a ground swell of negative publicity that may or may not be actually based on fact" and teaching that "the time lag between when the information is posted on the internet and what and how a company responds" must be minimized.). Thus, a company or brand would have been motivated to establish a rapid monitoring rate (e.g., with short intervals between the monitoring performed by the crawlers or bots) so that issues relating to the company or brand (e.g., a decline in sentiment) could be identified and addressed rapidly.

147. As noted above, Heath contemplates implementation of its platform in the form of "a system internal to" an organization." Ex. 1007, 49:4-5. A company implementing Heath's platform internally maps to "an entity" in the claim. Since the

platform internal to the company in Heath functions to identify (or select) the social media content sources for monitoring, the selecting of the data source “for monitoring” in Heath is performed by “the entity.”

148. Heath also teaches that “alerts can be sent for: extremely positive or negative posts; consumer or brand sentiment index changes; ... and posts containing specific keywords.” Ex. 1007, 64:67-65 (emphasis added). Heath does not expressly state that the keywords used by its monitoring function are the same “keywords expressing consumer or brand sentiment about the terms or phrases of interest” (*id.*, at 5:40-43) used by the forum analyzer for initially “scouring” forums, sub-forums and websites for the purpose of identifying relevant sources. However, it would have been obvious to use the same keywords for both the initial discovery of the “forums and/or sub-forums” that are “relevant to a specific user’s industry” and ongoing monitoring, since both processes are seeking to identify online activity relating to, e.g., a particular brand. *Compare, id.*, at 54:26-27 (discussing operation of initial forum analysis to identify “social media content (threads) relevant to a particular ... product/service”) *with, id.*, at 56:6-10 (discussing output of ongoing monitoring as including “sentiment towards a given brand, product or service.”) Thus, Heath teaches, or renders obvious, “monitoring for matches to the query ... when ... newly added data matches the query.”

e. Element 1.6 - “extracting data from the data source when at least ... the newly added data matches the query”

149. It would have been obvious for Heath’s monitoring to extract data from a monitored data source when new data matches the query in order to provide such information to a user in connection with an alert. Ex. 1007, 64:67-65 (discussing generation of alerts for "posts containing specific keywords.") Thus, Heath, alone or with Randall, teaches or renders obvious “extracting data from the data source when ... the update to the stored data matches the query.” See Ex. 1023, 3:7-12 (discussing deployment of web crawler system when links are "updated or changed.")

f. Element 1.7 - “determining an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period”

150. As explained above in Section XII(A)(1)(f), Laribi taught a “performance monitoring service” that monitored performance metrics of “data collection” applications as well as “data analysis” applications for the purpose of provisioning/deprovisioning one or more instances of an application. Ex. 1008, 1:31-46, 1:51-2:2, 2:7-10, 11:32-35, 13:7-24. Among other performance metrics, Laribi taught monitoring applications based on load. *Id.*, 2:19-23.

151. Heath discloses two interrelated processes, namely, the extraction of “raw” information and subsequent analysis of such information. Ex. 1007, 53:10-

23, 54:45-56:26 and 56:29-31. A POSITA would have understood that the extracted data in Heath requiring analysis corresponded to the load on Heath's subsequent data analysis process.

152. As explained below, in view of Laribi, a POSITA would have been motivated to monitor the load on Heath's data analysis process for the purpose of provisioning or deprovisioning a number of instances of Heath's data analysis application. A POSITA would have been motivated to provision additional instances of Heath's data analysis application at times when the rate of extracted data (or load) was high, so as to ensure that the extracted data was analyzed in a timely manner. Conversely, a POSITA would have been motivated to deprovision instances of the data analysis application at times when the rate of extracted data (or load) was low in order to conserve computer resources and/or power. Ex. 1008, 58:18-44 and 61:34-43.

153. As noted above, by 2013, corporations understood the need to respond rapidly to issues arising on social media relating to their brands. See, e.g., Ex. 1027, at [0021] (explaining that "before a company or entity is even aware of a problem, the viral nature of information spreading on the internet can lead to a ground swell of negative publicity that may or may not be actually based on fact" and teaching that "the time lag between when the information is posted on the internet and what and how a company responds" must be minimized.) Thus, a company or brand

would have been motivated to quickly analyze new messages being generated on social media so that issues relating to the company or brand (e.g., a rapid decline in sentiment) could be identified and addressed rapidly. More specifically, a POSITA would have been motivated to deploy Heath's platform in a manner that enabled rapid extraction and analysis of new messages generated on social media at times when the generation rate of such messages was high.

154. Given the motivation of companies to quickly respond to brand issues arising online, a POSITA would have been motivated to deploy Heath's system using a framework such as disclosed in Laribi, that allowed the platform to scale up its capacity to analyze social messages at times when the generation (and extraction) rate of such messages was high, and scale down such capacity when the message generation (and extraction) rate was slow in order to conserve computer resources and/or power, as taught in Laribi. Ex. 1008, 58:18-44 and 61:34-61.

155. Accordingly, it would have been obvious in view of Laribi to determine “an extraction rate for extracting the data, the extraction rate indicating an amount of the data that is extracted over a first time period” in Heath, so as to monitor the load on Heath’s subsequent data analysis process for the purpose of provisioning or deprovisioning a number of instances of Safford’s data analysis application.

g. Element 1.8 - “determining a first processing rate for processing the extracted data with a number of parallel processors, the first

**processing rate indicating an amount of
extracted data that is processed over a second
time period”**

156. In view of Laribi, it would have been obvious to monitor the data processing rate of Heath’s data analysis application and provision additional instances thereof to ensure that the data processing rate of Heath’s data analysis application was sufficient to timely process data being provided thereto by Heath’s monitoring process at times when the rate of content such as messages generated on social media was elevated. Accordingly, it would have been obvious in view of Laribi to determine “a first processing rate for processing the extracted data” in Heath where “the first processing rate” indicates “an amount of extracted data that is processed” over a given time.

157. As explained below in connection with element 1.9, the claimed “second time period” corresponds to a time when Heath’s data analysis application is performing at a processing rate (corresponding to the “first processing rate” in the claim) that is insufficient to timely analyze the load of extracted data being provided to Heath’s data analysis application by Heath’s extraction operation.

158. As explained above in Section XII(A)(1)(g), Laribi taught that multiple instances of an application could run on parallel processors. Accordingly, in view of Laribi, it would have been obvious to implement Heath’s data analysis process by deploying multiple instances of the application on parallel processors.

h. Element 1.9 “dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate to obtain a second processing rate that is greater than the first processing rate”

159. In view of Laribi, a POSITA would have been motivated to dynamically provision additional instances of Heath’s data analysis application on parallel processors at times when Heath’s data scraping application was performing at a processing rate (the “first processing rate”) that was insufficient to timely analyze the load of extracted data being provided to Heath’s data analysis application because the rate of content such as messages generated on social media was elevated. Subsequent to dynamically scaling up the number of instances of the data processing application based on such load (which maps to “dynamically adjusting the number of parallel processors for analyzing the extracted data based on the extraction rate”), the Heath/Laribi system will achieve a faster data analysis processing rate (or “second processing rate”) that is greater than “the first processing rate.”

i. Element 1.10 - “analyzing, with the parallel processors, the extracted data to determine at least one of a sentiment ... “

160. As noted in Sections XII(K)(1)(f)-(h), it would have been obvious to apply Laribi’s application performance monitoring and scaling techniques to the data analysis operation performed by Heath. Heath taught that such data analysis included determining a sentiment. Ex. 1007, 3:36-39 and 64:67-65:5.

j. Element 1.11 - “establishing a two-way communication channel, between at least the entity that selected the data source for monitoring and a user device of a user that provided the data to the data source, based on the analysis of the extracted data”

161. Heath taught that information from its data analysis operation (which is “based on the analysis of the extracted data”) may enable users “identify and target social media authors with the most influence to enter into online communications and activity with these lead authors and influence their opinion” Ex. 1007, 59:40-45; see also, *id.*, 56:6-10 (“processing also provides information including: the websites on which each of the social media authors have contributed; ... the status of influence of the authors; the author’s sentiment towards a given brand, product or service ...”) and *id.*, 64:67-65:5 (“alerts can be sent for: ... particular authors users wish to track”). Thus, Heath teaches identifying social media authors with which to engage based on the author’s influence and sentiment, which maps to identifying social media authors “based on the analysis of the extracted data.” In addition, such a social media author maps to “a user that provided the data to the data source.”

163. A POSITA would have understood that “enter[ing] into online communications” with “social media authors with the most influence” in order to “influence their opinion” in Heath would have included “establishing a two-way communication channel” between the company or brand and the social media author because such a communications channel would have been required to exchange

messages or posts between the company and the social media author on the social media platform hosting the social media. Social Media platforms like those discussed in Heath had direct messaging functionality that provided an established communication channel through which one user directly communicated with another user. Ex. 1007, 4:66-5:6; Ex. 1029, 14:39-43, Ex. 1030, [0100] and Ex. 1031, [0025]-[0026]. To the extent Heath is not explicit that such communications would take place over the social media platform, it would have been obvious for the company to communicate with the social media author over such channel because the social media author is readily accessible and engaged in social media.

164. As noted above, a company implementing Heath's platform internally to monitor its brand maps to "an entity" in the claim, and it is obvious that such an entity would also function to identify (or select) the social media content sources for monitoring. See, Sections XII(K)(1)(b) and (d). It would also have been obvious that the entity internally implementing Heath's platform for the purpose of monitoring the entity's brand would be the same entity that entered into online communications with "social media authors with the most influence" in order to "influence their opinion" about the entity's brand.

- k. **Element 1.12 - "transmitting, from the entity via the two-way communication channel, a first message directed to the user device based on the analysis of the extracted data" and**

Element 1.13 - “receiving, from the user device via the two-way communication channel, a second message in response to the first message directed to the user device”

165. A POSITA would have understood that “enter[ing] into online communications” with “social media authors with the most influence” in order to “influence their opinion” in Heath (see Section XII(K)(1)(j)) would have included transmitting messages to, and receiving responses from, the social media accounts of such authors.

166. Claim 1 is therefore obvious over Heath, Randall and Laribi.

2. Claims 11 and 16

167. The additional limitations of Claim 11 (“a memory unit” and “at least one processor ... configured” to perform the recited functions), and Claim 16 (“a non-transitory computer-readable medium having program code recorded thereon” to perform the recites functions), are met by Heath. Ex. 1007, 65:41-45 and 66:27-30. Thus, Claims 11 and 16 are obvious over Heath, Randall and Laribi.

3. Claims 2 and 12

168. Heath teaches a dashboard and graph that each visualize the extracted data. Ex. 1007, 63:65-64:2, 64:25-32 and Figures 22-23. Thus, Claims 2 and 12 are obvious over Heath, Randall and Laribi.

4. Claim 8 - “the monitoring comprises deploying extractors to the data source, a number of extractors based at least in part on a traffic volume of the data source”

169. As noted above, among other metrics used for scaling up or scaling down instances of an application, Laribi teaches monitoring metrics identifying a load across the plurality of instances of the application.” Ex. 1008, 2:18-20. In Heath, the load on the extraction process corresponds to the volume of traffic to be scraped. In view of Laribi, it would have been obvious to scale the number of instances of the Heath’s extraction application or Randall’s crawlers/robots (each instance of which maps to an extractor that monitors and extracts data from source(s) of interest) based on the volume of traffic, in order to dynamically adjust the number of instances of the application or crawlers/robots being deployed “based at least in part on a traffic volume of the data source” when, for example, the rate of such as messages generated on social media was elevated. A POSITA would have been motivated to dynamically adjust the number of instances of the application or crawlers/robots so as to match the capacity of number of instances to the current traffic load, in order to ensure timely extraction of the data. Thus, Claim 8 is obvious over Heath, Randall and Laribi.

5. Claims 9, 15 and 20

170. As explained in Section XII(K)(1)(i), Heath teaches or renders obvious use of social media as “the two-way communication channel.” Thus, Claims 9, 15 and 20 are obvious over Heath, Randall and Laribi.

6. Claim 10

171. As explained in Section XII(K)(1)(b), brand sentiment in Heath maps to “a topic of interest” that is analyzed based on the results of the query (keywords). Heath also analyzes “sentiment trends” which maps to a sub-topic of interest that is also analyzed based on the results of the query. Ex. 1007, 60:2-5.

172. As explained in Section XII(K)(1)(c), Heath determines a relevance score that reflects a correlation between a given social media source and a given brand based on keyword matches. Thus, Heath teaches determining a “correlation between the data source and the topic of interest based on the analysis of the results of the query,”

173. Accordingly, Claim 10 is obvious over Heath, Randall and Laribi.

7. Claim 19

174. As explained in Section XII(K)(1)(c), the forums, sub-forums and websites “scoured” by Heath’s “forum analyzer” map to “a plurality of data sources.” It would have been obvious to define a list of sources to be scoured before beginning the scouring process so that the scouring tasks could be efficiently divided in advance amongst the computer resources performing the probing. In such an

approach, the “plurality of data sources” would be “predefined.” Thus, Claim 19 is obvious over Heath, Randall and Laribi.

L. Ground 12 - Claims 3, 13 and 17 Are Obvious Over Heath, Randall, Laribi and Chen

175. As explained in Sections XII(B), the additional limitations of Claims 3, 13 and 17 are met by Chen. Heath and Chen are both directed to methods for monitoring online activity pertaining to a particular brand. A POSITA would have motivated to apply Chen's query refinement techniques to Heath's keywords in order to reduce irrelevant information identified by the keywords, and would have had a reasonable expectation of success with respect to the combination. Thus, Claims 3, 13 and 17 are obvious over Heath, Randall, Laribi and Chen.

M. Ground 13 - Claims 4, 14 and 18 Obvious Over Heath, Randall and Laribi In View Of Seibel

176. As explained in Sections XII(C), the additional limitations of Claims 4, 14 and 18 are met by Seibel. It would have been obvious to apply Heath's techniques for determining sentiment to a combination of social media channels (“open-source data”) and internal corporate documents “such as emails” and “customer support notes” (“closed-source data sources”) as taught in Siebel, in order to expand the scope of data being analyzed and more accurately quantify sentiment based on a combination of social media posts from a particular user (e.g., an advocate) and internal corporate documents “such as emails” and “customer support notes”

representing private communications between the company and such user. Thus, Claims 4, 14 and 18 are obvious over Heath, Randall, Laribi and Seibel.

N. Ground 14 - Claims 6-7 Are Obvious Over Heath, Randall, Laribi, Schultz and Gravano

177. It would have been obvious to apply Schultz's query expansion techniques to Heath's keywords in order to generate a more robust cluster of keywords. Query expansion was a well-known technique that was beneficial, e.g., in cases where it is difficult to use a single term to represent an information concept. In a case where a single search term was insufficient to represent the concept being queried, a POSITA would have been motivated to apply query expansion techniques in order to more fully represent the concept being queried within the keywords. Moreover, for the same reasons discussed in Section XII(D)(1) in connection with Safford, a precision-oriented company wishing to monitor its brand using Heath's platform would have been motivated to identify for monitoring social media accounts having a high value of *P* in order to lessen the proportion of useless items from the social media accounts being monitored, as taught in Gravano. Thus, Claim 6 would have been obvious over Heath, Randall, Laribi, Schultz and Gravano.

178. Claim 7 depends from Claim 6. As explained in Section XII(D)(2), the additional limitation of Claim 7 was obvious over Schultz. Thus, Claim 7 would have been obvious over Heath, Randall, Laribi, Schultz and Gravano.

O. Ground 15 - Claim 5 Is Obvious Over Heath, Randall and Laribi, In View Of Schultz and Moulinier

179. As explained in Section XII(E), the additional limitations of Claim 5 were met by Schultz and Moulinier. Heath, Schultz and Moulinier are all directed to search query technology. It would have been obvious to apply Moulinier's language translation techniques to Schultz's natural language query. In a case where Heath's social media accounts were in different languages, a POSITA would have been motivated to use Moulinier's language translation techniques with Schultz/ Heath, so as to permit the user to search social media accounts in different language based on a natural language query in the user's native language. Thus, Claim 5 is obvious over Heath, Randall and Laribi, in view of Schultz and Moulinier.

XIII. DECLARATION

180. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of tile 18 of the United States Code.

Dated: March 4, 2025



Dr. Jaideep Srivastava, Ph.D.