

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

BRIGHT DATA LTD.,
Petitioner,

v.

OXYLABS, UAB,
Patent Owner.¹

IPR2023-01441
Patent 7,010,526 B2

Before THOMAS L. GIANNETTI, STACEY G. WHITE, and
IFTIKHAR AHMED, *Administrative Patent Judges*.

WHITE, *Administrative Patent Judge*.

DECISION
Denying Institution of *Inter Partes* Review
35 U.S.C. § 314

¹ The caption of this proceeding has been revised to reflect the current ownership of the challenged patent. *See infra*, Section I, n.2.

I. INTRODUCTION

Bright Data Ltd. (“Petitioner”) filed a Petition requesting an *inter partes* review of claims 1–3, 6–9, 11, 13, 16, 17, 19, 20, 34–36, 39–42, 44, 46, 49, 50, 52, and 53 of U.S. Patent No. 7,010,526 B2 (Ex. 1001, “the ’526 patent”). Paper 2 (“Pet.”). Oxylabs, UAB² (“Patent Owner”) filed a Preliminary Response. Paper 6 (“Prelim. Resp.”).

Under 35 U.S.C. § 314(a), an *inter partes* review may not be instituted unless the information presented in the petition “shows that there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition.”

For the reasons stated below, we determine that Petitioner has not established a reasonable likelihood that it would prevail with respect to at least one claim. We, therefore, deny institution of *inter partes* review.

A. Related Matters

The parties indicate that the ’526 patent is involved in the following district court proceeding: *Bright Data Ltd. v. Oxylabs, UAB*, No. 2:23-cv-171 (E.D. Tex.).³ Pet. 1; Paper 4, 1; Prelim. Resp. 16.

² Oxylabs, UAB is identified on the face of the petition as Patent Owner and Patent Owner’s Mandatory Notices state that oxylabs, UAB is the current owner. Pet, Paper 4, 1. Patent Owner’s Preliminary Response, however, identifies Metacluster LT, UAB as Patent Owner on the face of that document. Prelim. Resp. Based on the representation of ownership provided in the Mandatory Notices, we will refer to oxylabs, UAB as Patent Owner. If this is incorrect, the parties are to contact the Board within 14 days of this Decision to schedule a conference call.

³ Patent Owner identifies this case as *Metacluster LT, UAB v. Bright Data Ltd.*, No. 2:23-cv-00171 (E.D. Tex.). Paper 4, 1.

B. The '526 Patent

The '526 patent is titled “Knowledge-Based Data Mining System,” and describes a data mining system employing data miners. Ex. 1001, codes (54), (57). “Data mining is the process of extracting user-desired information from a corpus of information. Perhaps the most widespread example of data mining is the search engine capability incorporated into most Web browsers.” *Id.* at 1:11–14. Prior art search engines created a relevancy ranking by indexing the words that occur on a web page and correlating those words with the page’s uniform resource locator (“URL”). *Id.* at 1:19–36. The '526 patent explains that, “[d]espite the relevancy ranking used by most commercial search engines, finding particular types of information typically entails a great deal of mundane sifting through query results by a person.” *Id.* This was true because expertise was often necessary to determine which pages were the most useful and relevant. *Id.* at 1:35–37.

The '526 patent purports to address these problems with a system and method for data mining that includes crawling the Internet to return web pages to a data store, and then accessing the data store with at least a first data miner to process data in the store and output at least one result. *Id.* at 2:10–19. The result output by the first data miner may include one or more keys representing one or more characteristics of the data. *Id.* at 2:33–38. For example, an entity may be a web page and a key may represent whether that web page represents a corporation. *Id.* at 2:40–47. Upon receiving a customer request for information, at least a second data miner is created in response to the customer request for information, with the second data miner accessing at least the result of the first data miner to generate an output. *Id.*

at 2:10–19. Based on the output of the second data miner, the customer request is fulfilled. *Id.* Data miners use expert rules to process entities and append respective keys to the entities representing characteristics of the entities as derived from expert rules embodied in the miners. *Id.* at code (57). With these keys, characteristics of entities as defined by disparate expert authors of the data miners are identified for use in responding to complex data requests from customers. *Id.*

Figure 1 of the '526 patent is reproduced below.

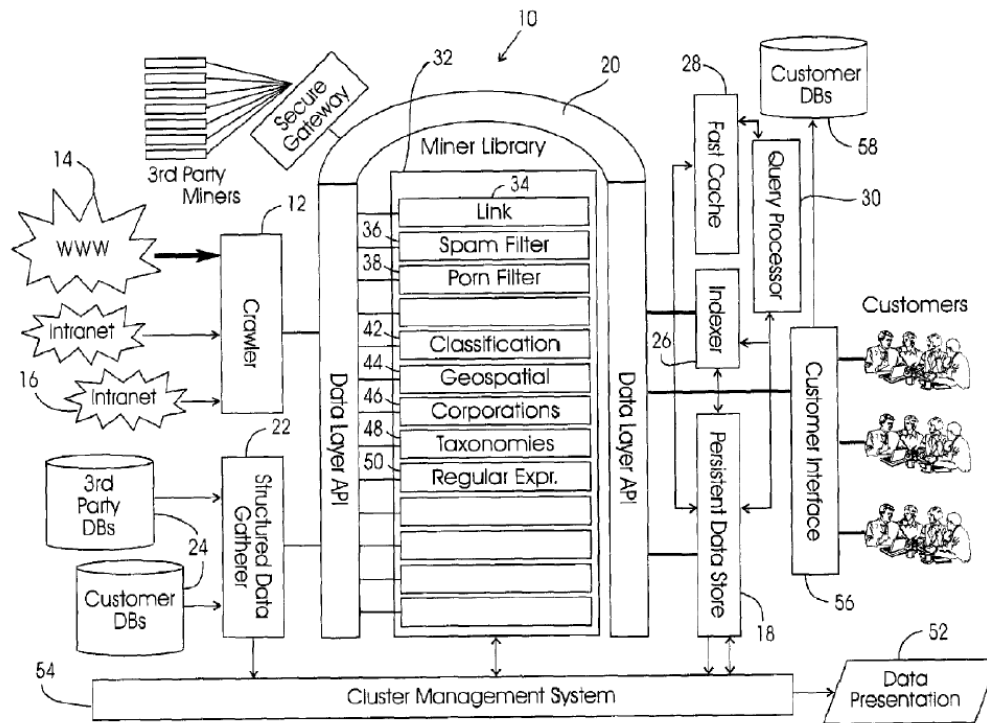


FIG 1

Figure 1 illustrates a data mining system 10 for responding to customer requests for data. *Id.* at 2:64–65, 3:7–15.

“[S]ystem 10 essentially integrates the knowledge of many experts to sift through a large corpus of data to respond to what might be relatively complex requests for information.” *Id.* at 3:9–11. System 10 can be used for enterprise data analysis, competitive intelligence, trending, discovery,

web portal services, clustering, and taxonomy creation. *Id.* at 3:7–24.

System 10 includes a data gathering layer, a data storage layer, a data mining layer, a data presentation layer, and a system management layer. *Id.* at 3:25–43. In the data gathering layer, Web Crawler 12 accesses World Wide Web 14. *Id.* Crawler 12 continuously crawls World Wide Web 14 and outputs the crawled pages to data store 18 using a data layer application programming interface (“API”) 20. *Id.*

Data store 18 contains a relatively large corpus of data, such as web page data from crawler 12. *Id.* at 4:16–30. Data store 18 also contains entities that represent the underlying data. *Id.* These entities have respective universal identities (UEIDs) that encode the entity identification and entity type, such as “web page”, “hyperlink”, “person”, “corporation”, or “article.” *Id.* Entities can also contain keys with associated key values as appended to the entities by the data miners. *Id.* Indexer 26 includes indices of keys and key values found in the store. *Id.* at 4:31–38

The ’526 patent provides that “[b]y ‘miner’ or ‘data mining element’ is meant an analysis engine that generates an output, and specifically an output that can include one or more keys representing characteristics of an entity, using a set of rules.” *Id.* at 5:63–6:1. The rules may be heuristically determined, and may include statistically-based rules. *Id.* Data miners are modular components that have specific input and output specifications. *Id.* at 6:25–40. Miners are specified in terms of the data, usually as indicated by keys, that must be available when the miners start, and the data (including other keys) that the miners will create during successful processing. *Id.* A miner may consume work from a system-managed work queue, based on one or more dependencies that are specified by the miner. *Id.* at 6:41–45.

Miners read long-running, persistent, and reliable streams of raw content, as well as processed data created by other miners, from data store 18. *Id.* at 7:4–27.

Miners write back the results of their processing to data store 18, for other miners and end users to access. *Id.* at 7:28–33. To write data back into data store 18 for other miners to access, a miner creates new keys and values to attach to the entity, then performs a store write operation. *Id.* The results of a particular customer request for information as provided by the miners is then presented on data presentation layer 52. *Id.* at 7:34–38.

Figure 2 of the '526 patent is reproduced below.

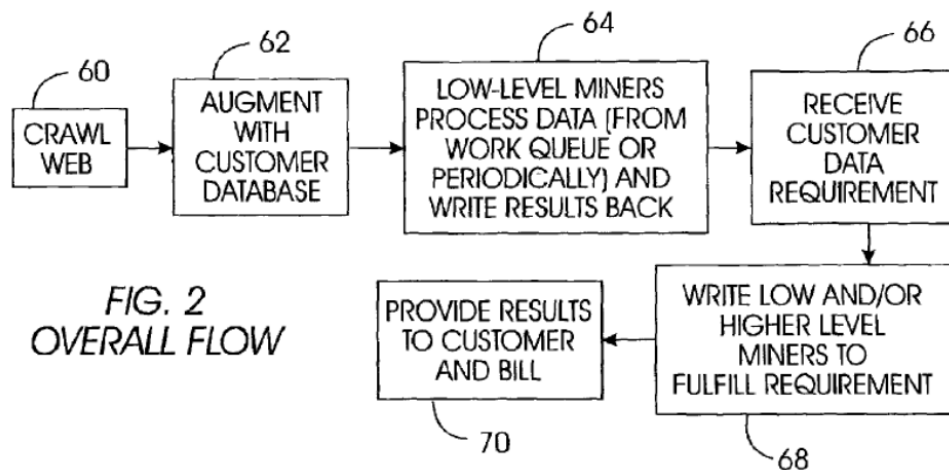


Figure 2 illustrates the overall logic of the operation of data mining system 10. *Id.* at 2:66, 8:13–15.

The logic of data mining system 10 starts with block 60 at which crawler 12 crawls the Web to add data to data store 18. *Id.* at 8:13–31. At block 62, data store 18 may be augmented with data from databases 24, using data gatherer 22. *Id.* Once data store 18 contains data, the logic flows to block 64, where at least some of the miners—which are considered “low level” miners—access data and process it, before writing the results back to data store 18. *Id.* For example, filtering miners (such as a SPAM filter) may

process all Web pages in data store 18 and write back respective keys to the corresponding entities indicating whether each site is SPAM. *Id.*

Next, at block 66, customer requests for information are received. *Id.* at 8:32–40. At block 68, additional low level miners may be written in response, or high level miners, if required and not yet written, may be created. *Id.* High level miners “can be thought of as miners that specify dependencies on the outputs of other miners, i.e., that require entities for processing that have been tagged with keys output by lower level miners.” *Id.* An example of a higher level miner is one that responds to the query, “what do our commercial customers in the Pacific Northwest think of our competitor’s health care products in terms of brand name strength and value?” *Id.* at 8:40–56. Such a miner may specify that it wishes to receive only pages from the Pacific Northwest (as indicated by a geospatial key appended to entities by the geospatial miner), and only if the competitor’s name is featured in the entity (as indicated by a key appended to the entity by a proper name miner). *Id.* According to the ’526 patent, “[m]any such dependencies might be hypothesized, it being understood that the expert who might specify the dependencies of such a miner uses heuristics according to his or her expertise without having to know how the expert who wrote, e.g., the geospatial miner arrived at his or her solution.” *Id.* The results are finally provided to the customer at block 70. *Id.*

C. Illustrative Claim

Challenged claims 1 and 34 are independent. Challenged claims 2, 3, 6–9, 11, 13, 16, 17, 19, and 20 depend from claim 1, and challenged claims 35, 36, 39–42, 44, 46, 49, 50, 52, and 53 depend from claim 34. Claim 1 is illustrative of the subject matter of the challenged claims.

1. A method for data mining, comprising:
 - crawling the Internet to return Web pages to a data store;
 - accessing the data store with at least a first data miner to process data in the store and output at least one result;
 - receiving a customer request for information;
 - creating at least a second data miner in response to the customer request for information, the second data miner accessing at least the result of the first data miner to generate an output; and
 - at least partially based on the output of the second data miner, responding to the customer request for information.

Ex. 1001, 11:63–12:7.

D. Asserted Grounds of Unpatentability

Petitioner, supported by the declaration of Tim A. Williams, Ph.D. (Ex. 1008), asserts the following grounds of unpatentability (Pet. 4)⁴:

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
1, 3, 19, 20, 34, 36, 52, 53	§ 102	Jacobsen ⁵
2, 35	§ 103	Jacobsen
2, 35	§ 103	Jacobsen, Jaramillo ⁶
6–9, 11, 16, 17, 39–42, 44, 49, 50	§ 103	Jacobsen, Hegli ⁷

⁴ We apply the pre-AIA version of 35 U.S.C. § 102 and § 103 because the claims at issue have an effective filing date prior to March 16, 2013, the effective date of the Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) (“AIA”). See Ex. 1001, code (22).

⁵ Jacobsen et al., U.S. Pat. Appl. Pub. 2003/0163454 A1, published Aug. 28, 2003 (Ex. 1003, “Jacobsen”).

⁶ Jaramillo, U.S. Pat. Appl. Pub. 2002/0049773 A1, published Apr. 25, 2002 (Ex. 1005, “Jaramillo”).

⁷ Hegli et al., U.S. Pat. No. 6,606,659 B1, issued Aug. 12, 2003 (Ex. 1006, “Hegli”).

Claim(s) Challenged	35 U.S.C. §	Reference(s)/Basis
1, 6, 9, 11, 13, 19, 20, 34, 39, 42, 44, 46, 52, 53	§ 102	Seibel ⁸
2, 35	§ 103	Seibel

In response, Patent Owner relies on the declaration of Kevin C. Almeroth, Ph. D. (Ex. 2001).

II. ANALYSIS

A. Level of Ordinary Skill in the Art

In order to determine whether an invention would have been obvious at the time the application was filed, we consider the level of ordinary skill in the pertinent art at critical time. *Graham v. John Deere Co.*, 383 U.S. 1, 17 (1966). The resolution of this question is important because it allows us to “maintain[] objectivity in the obviousness inquiry.” *Ryko Mfg. Co. v. Nu-Star, Inc.*, 950 F.2d 714, 718 (Fed. Cir. 1991). In assessing the level of ordinary skill in the art, various factors may be considered, including the “type of problems encountered in the art; prior art solutions to those problems; rapidity with which innovations are made; sophistication of the technology; and educational level of active workers in the field.” *In re GPAC, Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995) (quotation omitted). Generally, it is easier to establish obviousness under a higher level of ordinary skill in the art. *Innovation Toys, LLC v. MGA Entm’t, Inc.*, 637 F.3d 1314, 1323 (Fed. Cir. 2011) (“A less sophisticated level of skill generally favors a determination of nonobviousness . . . while a higher level of skill favors the reverse.”).

⁸ Seibel et al., U.S. Pat. No. 7,082,427 B1, issued July 25, 2006 (Ex. 1004, “Seibel”).

Petitioner asserts that a person having ordinary skill in the art at and before the priority date of the '526 patent “would have had . . . either (1) a Bachelor’s Degree or higher in the field of electrical engineering, computer engineering, or computer science and at least one year of related practical experience; or (2) at least four years of practical experience in data management.” Pet. 13 (citing Ex. 1008 ¶¶ 30–31). At this stage of the proceeding, Patent Owner does not dispute the level of ordinary skill in the art. *See* Ex. 2001 ¶ 35.

We are persuaded that Petitioner’s description is consistent with the prior art and patent specification before us. *See Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001). For the purpose of our Decision, we adopt Petitioner’s description.

B. Claim Construction

In an *inter partes* review proceeding, a patent claim shall be construed using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. § 282(b). 37 C.F.R. § 42.100(b). Our rule adopts the same claim construction standard used by Article III federal courts, which follow *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc), and its progeny. Under this standard, the words of a claim are generally given their “ordinary and customary meaning,” which is the meaning the term would have to a person of ordinary skill at the time of the invention, in the context of the entire patent including the specification. *See Phillips*, 415 F.3d at 1312–13.

The parties argue the construction of the phrase “creating at least a second data miner in response to the customer request for information,” which is recited in both of the challenged independent claims. Pet. 13–14;

Prelim. Resp. 7–10. Petitioner submits that a person of ordinary skill in the art “would understand the limitations of claims 1 and 34 to require creating a **new** data miner to search the results of a first data miner to provide customized results to the user based on the customized parameters of the user search request.” Pet. 14 (emphasis in original) (citing Ex. 1008 ¶ 49). Petitioner goes on to state that “[t]his understanding is consistent with the Merriam-Webster dictionary definition of ‘creating’ as ‘to make or bring into existence something new.’” *Id.* (citing Ex. 1007, 2).

Patent Owner responds that it “does not dispute the construction of a ‘new data miner’ at this point, but this is not the only modification that Petitioner makes to the claim.” Prelim. Resp. 8. Patent Owner contends that Petitioner’s construction overlooks the claim language “in response to the customer request for information.” *Id.* According to Patent Owner, “[t]here is no logical connection between the proposed definition of creating and Petitioner’s rewording of the claim to replace the feature ‘in response to the customer request for information’ with ‘based on the customized parameters of the user search request.’” *Id.* at 9.

Based on the record before us, we find Patent Owner’s argument to be persuasive. As both parties note, the phrase in question recites “creating” a second data miner, which requires the generation of a new data miner. Petitioner’s declarant testifies that during prosecution of the ’526 patent prior art references were “expressly distinguished” because the prior art merely “pick[ed] a pre-existing search engine” or fell short because “no new miners [were] created.” Ex. 1008 ¶ 49 (citing Ex. 1002, 83, 34); *accord* Ex. 2001 ¶ 57 (similar testimony from Patent Owner’s declarant). We agree with the parties that this phrase requires the generation of a new data miner.

Patent Owner argues that Petitioner's construction falls short because it does not properly address the requirement that the second data miner be created "in response to the customer request." Patent Owner states, and we agree, that "provid[ing] customized results to the user based on the customized parameters of the user search request" does not address the requirement that the second data miner be created "in response to" a customer request. Prelim. Resp. 9. In support of Patent Owner's assertions, Patent Owner's declarant points out that "as described with reference to FIG. 2 of the '526 Patent, 'customer requests for information can be received' at block 66 and then at block 68, [']additional low level miners can be written in response, or high level miners, **if required and not yet written, can be created.**" Ex. 2001 ¶ 61 (citing Ex. 1001, 8:32–35) (emphasis added).

As stated in the '526 patent a "'miner' or 'data mining element' is . . . an analysis engine that generates an output, and specifically an output that can include one or more keys representing characteristics of an entity, using a set of rules." Ex. 1001, 5:63–66. The '526 patent further notes that "the data miners are modular components that have specific input and output specifications." *Id.* at 6:25–26. Further, "[m]iners are specified in terms of the data, usually as indicated by the below-described keys, that must be available when they start, and the data (including other keys) that they will create during successful processing." *Id.* at 6:36–40. Thus, our review of the intrinsic record persuades us that creation of a second data miner requires the creation of a new "analysis engine," and not just the provision of customized parameters to an existing "analysis engine." Therefore, for the purposes of this decision, we construe the phrase "creating at least a second data miner in response to the customer request for information," to

require (1) the creation of at least one new second data miner, (2) that this creation be done in response to a customer request.

For purposes of this Decision, and based on the record before us, we determine that no other express construction of any other term is necessary. *See Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) (“[W]e need only construe terms ‘that are in controversy, and only to the extent necessary to resolve the controversy.’”).

C. Overview of the Asserted Prior Art

1. Jacobsen (Ex. 1003)

Jacobsen is titled “Subject Specific Search Engine” and relates to “a search engine tailored to search the Internet and to return results that contain fewer irrelevant results than present search engines return.” Ex. 1003, code (54), ¶ 2. Jacobsen’s subject-specific search engine (“SSSE”) utilizes a smart web crawler and includes a capability of filtering out sites not relevant to a particular subject. *Id.* at code (57), ¶ 10

Jacobsen’s Figure 4 is reproduced below.

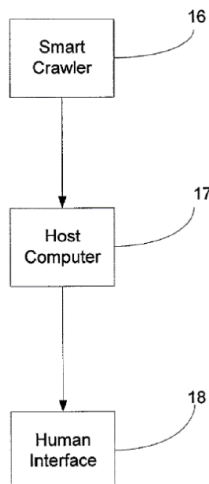


Figure 4

Figure 4 illustrates a configuration of an SSSE. *Id.* ¶¶ 39, 44.

The SSSE illustrated in Figure 4 provides a subject matter specific search engine that returns only links to resources that contain terminology relevant to that subject. *Id.* ¶ 19. Jacobsen explains that the operation of its SSSE is based on the principle that “most pages are not relevant,” and the “SSSE should separate the ‘straw’ from the ‘chaff.’” *Id.* ¶ 75. The SSSE, thus, does not visit every page on the Internet because the SSSE can determine quickly that a site is not relevant and, as a result, all the pages on that site are not indexed. *Id.* The system is able to build a fairly large database of relevant material very rapidly. *Id.*

The SSSE includes three primary components: smart crawler 16, which operates in a manner similar to prior art crawlers, i.e., by visiting sites and transmitting information back to host computer 17; host computer 17; and human interface 18, which is provided for entering search queries, and for human interaction in the processes of information screening and indexing. *Id.* ¶¶ 44, 45. Unlike prior art crawlers, smart crawler 16 does not operate under the “everything is relevant” principle. *Id.* ¶ 45. Smart crawler 16 traverses the Internet and performs a screening operation which determines, based on terminology, whether each site is relevant to the subject. *Id.* ¶¶ 25, 45. The result of using the site filter is that some sites are filtered out, leaving, essentially, a portion of the Internet containing only relevant sites. *Id.* ¶ 45. Only the information on relevant sites is transmitted to host computer 17, where the information on relevant sites, may be stored in a memory for further processing, or may be indexed and stored in a database. *Id.*

Jacobsen’s Figure 6A is reproduced below.

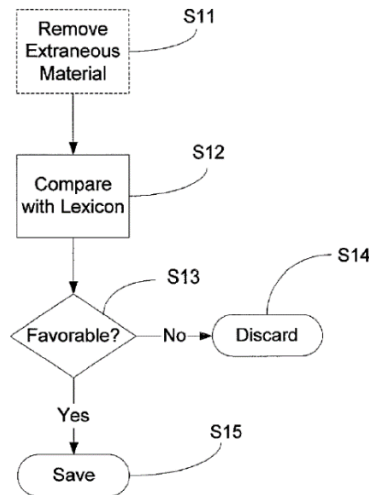


Figure 6A

Figure 6A depicts automated filtering operations in an SSSE. *Id.* ¶¶ 41, 61.

As depicted in Figure 6A, a web page is input to the filter, and (at optional step S11) extraneous material (e.g., advertisement) is removed. *Id.* ¶ 61.

The filter compares the page with a lexicon of terms at step S12, and the presence of such terms will indicate that the page may be relevant. *Id.* If the comparison is favorable (step S13), then the page is saved (step S15); if the comparison is unfavorable, the page is discarded (step S14). *Id.* One advantage of using a lexicon-based filter is that all of the components of the filter may be the same for any context/profession (except for the lexicon), and one need only change the lexicon accessed by the other components in order to create a search engine for a different context or profession. *Id.* ¶ 66.

A site ranking feature also may be included in the SSSE. *Id.* ¶ 67. The ranking system analyzes the relevant sites to determine the degree to which sites have been found interesting by others in the desired context. *Id.* This ranking may be performed by finding the number of links and citations to sites from other relevant sites. *Id.* Site ranking may be implemented via a

word ranking scheme. *Id.* ¶ 68. This technique assigns numerical scores to words and sums the scores of the words on a page to determine a score for the page. *Id.* The technique works by examining each non-trivial word (i.e., not “stop words,” like “if,” “it,” “and,” and the like) on a given page and increasing its score if it appeared on a relevant page (i.e., a page that passed through filtering) containing a link to the given page. *Id.* The word score is increased according to how many relevant pages that linked to the given page contain the word. *Id.* The technique may be augmented by increasing a word’s score according to where it appears in a link leading to the page being examined; for example, if the word appears closer in proximity to the link to the page being examined, its score is increased. *Id.*

A word score is saved for each word on each page. *Id.* ¶ 69. When a user enters a query, the SSSE determines a set of relevant pages that contain the query terms, and for each page, the word scores are summed to compute a site ranking for that page. *Id.* The site rankings for the pages then are used in determining how to display the search results to the user. *Id.* Thus, the SSSE utilizes dynamic site rankings, computed based on word rankings and in response to user queries. *Id.*

Search results in response to a user’s query may be displayed in a user-friendly manner. *Id.* ¶ 70. The most relevant sites, as determined by their site rankings, are displayed earlier in the display and/or more prominently than less high ranking sites. *Id.*

2. *Seibel (Ex. 1004)*

Seibel relates to “business intelligence software tools for acquiring leads for prospective customers, using Internet data sources.” Ex. 1004,

1:30–33. Seibel’s text indexing system is used in a lead generation system accessible by the client via the Internet. *Id.* at 2:9–11.

Figure 2 of Seibel is reproduced below.

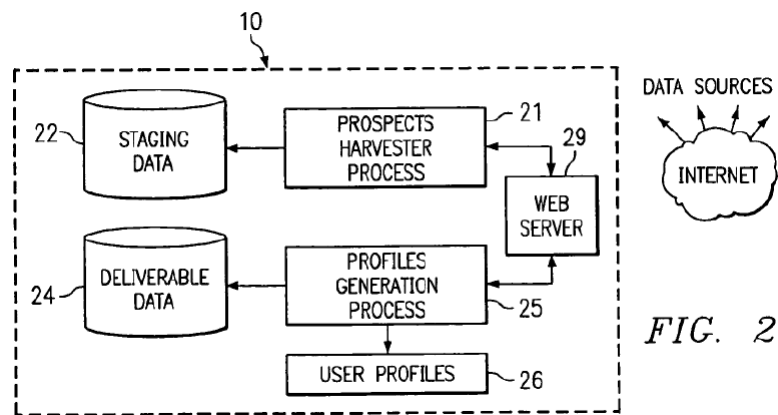


Figure 2 illustrates the functional elements of a lead generator system 10. *Id.* at 2:25–27, 3:58–62.

Seibel explains that the “functions of [lead generator] system 10 [shown in Figure 2] are partitioned between two distinct processes.” *Id.* at 3:59–62. The processes are prospects harvester process 21 and profiles generation process 25. *Id.* at Fig. 2.

Prospects harvester process 21 uses a combination of external data sources, client internal data sources, and user-parameter extraction interfaces, in conjunction with a search, recognition and retrieval system, to harvest contact information from the web, and return it to staging database 22. *Id.* at 3:63–4:5. Process 21 collects business intelligence data from both inside a client’s organization and outside the organization. *Id.* The information collected can be structured data (as in corporate databases/spreadsheet files) or unstructured data (as in textual files). *Id.* Process 21 may be programmed to validate and enhance data, utilizing a system of lookup, reverse lookup, and comparative methodologies that maximize the value of the contact information. *Id.* at 4:6–22.

Profiles generation process 25 analyzes the user's in-house files and records related to the user's existing customers to identify and group those customers into profile categories based on the customer's buying patterns and purchasing volumes. *Id.* at 4:23–46. The patterns and purchasing volumes of existing customers are overlaid on the salient contact information previously harvested, to allow the aggregation of the revenue-based leads into prioritized demand generation sets. *Id.* Process 25 uses an analysis engine and both data and text mining engines, to mine a company's internal client records, digital voice records, accounting records, contact management information, and other internal files. *Id.* Process 25 creates a profile of the most profitable customers, reveals additional prospecting opportunities, and enables sales cycle improvements. *Id.* The resulting profiles are then overlaid on the data obtained by process 21, to facilitate more accurate revenue projections and to enhance the sales and marketing process. *Id.* Results are stored in deliverable database 24. *Id.* Profiles generation process 25 may be used to create a user (client) profiles database 26, which stores profiles of the client and its customers. *Id.* at 4:47–52. Database 26 may be accessed during various data and text mining processes to better identify prospective customers of the client. *Id.*

Web server 29 provides the interface between client systems and lead generation system 10. *Id.* at 4:53–60. Web server 29 may route different types of requests to different sub processes within system 10. *Id.*

Figures 4 and 5 of Seibel, reproduced below, illustrate a database server system that is used within the lead generation system (such as that shown in Figure 2). *Id.* at 2:29–31.

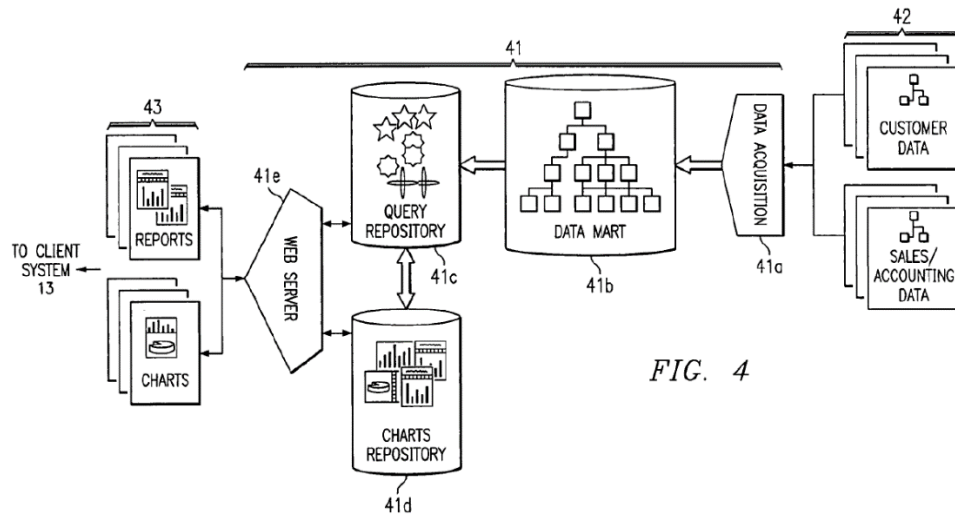


FIG. 4

Figure 4 illustrates elements of a database server system 41 used within lead generation system 10. *Id.* at 2:29–31, 5:59–62.

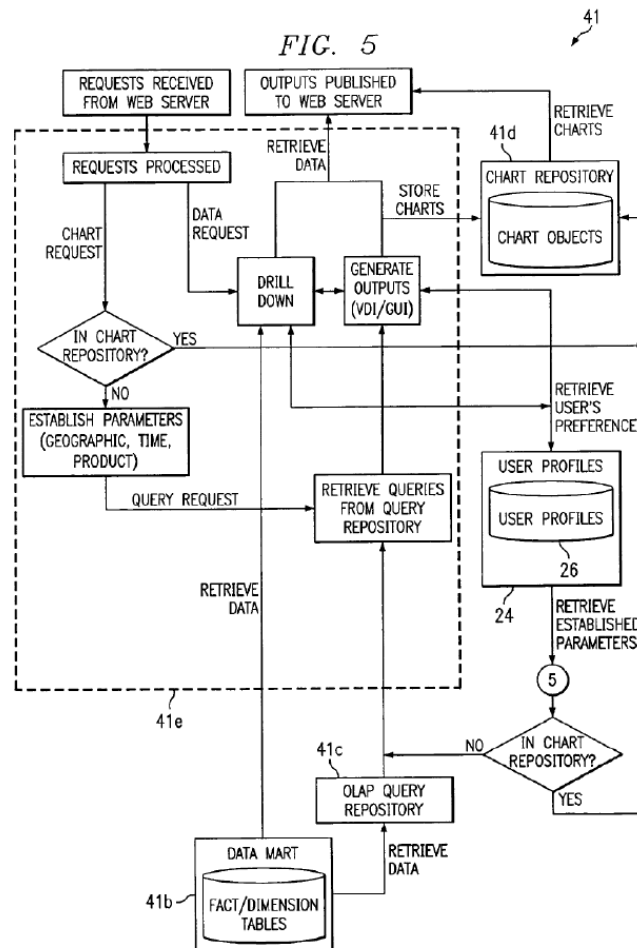


FIG. 5

Figure 5 is a data flow diagram illustrating data flow with respect to database server system 41 illustrated in Figure 4. *Id.* at 2:29–31, 5:59–62.

Database server system 41 may be used to implement profiles generation process 25, which collects profile data about a client. *Id.* at 5:59–64. Input data 42 (shown in Figure 4) may be the client’s sales data, customer-contact data, customer purchase data, account data, etc. *Id.* at 5:65–6:25. Database server system 41 has five major components: (1) Acquisition Component 41a, which collects and extracts data from different data sources, applying appropriate transformation, aggregation and cleansing to the data collected; (2) Data Mart 41b, which captures and stores an enterprise’s sales information; (3) Predefined Query Repository Component 41c, which is the central storage for predefined queries; (4) Business Intelligence Charts Repository Component 41d; and (5) Web Server component 41e. *Id.* at 6:10–26, 6:44–46, 6:58–61, 7:11–15.

Acquisition component 41a consists of predefined data conversions to accomplish commonly used data transformations. *Id.* at 6:10–25. The sales data collected from data acquisition component 41a are “sliced and diced” into multidimensional tables by time, region, product and customer dimensions, etc. *Id.* at 6:25–43. The total number of dimension tables and fact tables varies from customer to customer, and data mart 41b is designed to accommodate data collected from a majority of commonly used software packages. *Id.* More particularly, data mart 41b stores sales and accounting fact and dimension tables that will accommodate the data extracted from a majority of industry accounting and customer contact software packages. *Id.*

Predefined queries stored in Predefined Query Repository Component 41c are parameterized macros/business rules that extract information from tables in data mart 41b. *Id.* at 6:44–57. The results of these queries are delivered as business charts in a web browser environment to the end users.

Id. As an example of operation of database server system 41, gross revenue analysis of worldwide sales may be contained in predefined queries that are stored in query repository 41c. *Id.* at 7:40–46.

Business Intelligence Charts Repository Component 41d captures and stores the most frequently visited charts in a central location. *Id.* When an end user requests a chart, system 41 first queries chart repository 41d to see if there is an existing chart. *Id.* If there is a preexisting chart, server 41e pulls that chart directly from the repository. *Id.* If there is no preexisting chart, server 41e runs the corresponding predefined query from query repository 41c in order to extract data from data mart 41b and subsequently feed the data to the requested chart. *Id.*

Web Server component 41e has a number of subcomponents. *Id.* at 7:11–28. A web server subcomponent serves HTTP requests. *Id.* A database server subcomponent provides database drill-down functionality. *Id.* An application server subcomponent routes different information requests to different other servers. *Id.* For example, sales revenue chart requests will be routed to database system 41, customer profile requests will be routed to a Data Mining server, and competition information requests will be routed to a Text Mining server. *Id.* Another subcomponent of server 41e is a chart server, which receives requests from the application server, and either runs queries against data mart 41b using query repository 41c, or retrieves charts from chart repository 41c. *Id.*

Database server system 41 delivers output 43, which includes business intelligence about an organization's sales performance, as charts over the Internet or corporate Intranet. *Id.* at 7:29–39. Users can pick and choose

charts by regions, by quarters, by products, by companies, and by different chart styles. *Id.*

3. *Jaramillo (Ex. 1005)*

Jaramillo is titled “Method of Providing a Reference Standard Within a Web Site on an Internet.” Ex. 1005, code (54). Jaramillo’s method includes the steps of (i) providing a web site having a reference standard database on a computer network, (ii) identifying if a user has permission to access the reference standard database on the web site and selecting a search criteria for a reference standard by the user, if the user has permission to access the reference standard database, and (iii) searching the reference standard database for the reference standard using the search criteria and providing the user with the reference standard identified by the search of the reference standard database. *Id.* ¶ 12. Jaramillo describes charging a user to access a particular web site and for a user to pay a fee for the service provided by the web site on a per-use basis. *Id.* ¶ 26.

4. *Hegli (Ex. 1006)*

Hegli is titled “System and Method for Controlling Access to Internet Sites,” and “relates to systems and methods for selectively blocking access to particular Internet websites and pages” and “to a flexible filtering system and method that provides administrators with several options for controlling end-user access to those sites.” Ex. 1006, code (54), 1:8–14. Hegli’s system includes a database of Internet sites that have been categorized based on the type of information contained therein. *Id.* at code (57). Users are allowed to access sites within a particular category a limited number of times. *Id.* Users may request postponed access, wherein the site requested is stored at a server, and is available at a later time. *Id.* In addition, if a user chooses to

access a site within certain predefined categories, they are presented with the option of retrieving the page, but notified that their access will be logged to a file. *Id.*

D. Alleged Anticipation of Claims 1, 3, 19, 20, 34, 36, 52, and 53 By Jacobsen

Petitioner alleges that claims 1, 3, 19, 20, 34, 36, 52, and 53 were anticipated by Jacobsen, citing the Declaration of Dr. Williams for support. Pet. 4, 22–32 (citing Ex. 1008). Patent Owner opposes, citing the Declaration of Dr. Almeroth for support. Prelim. Resp. 28–35 (citing Ex. 2001).

For reasons discussed below, we determine that Petitioner has not shown a reasonable likelihood that it would prevail in its anticipation challenge based on Jacobsen.

1. Independent Claims 1 and 34

Independent claim 1 and 34 both recite “creating at least a second data miner in response to the customer request for information, the second data miner accessing at least the result of the first data miner to generate an output.” Ex. 1001, 12:1–4 (claim 1), 13:48–51 (claim 34). Petitioner relies on the same disclosures from Jacobsen as to these limitations from claims 1 and 34. Pet. 30. Thus, this discussion applies to Petitioner’s assertions regarding both of the challenged independent claims.

Petitioner argues Jacobsen discloses the claimed creation of a second data miner through its discussion of Jacobsen’s site ranking feature. Pet. 26. Petitioner asserts that Jacobsen’s site ranking feature accesses the word scores, which are alleged to be the output of the first data miner. *Id.* According to Petitioner, “[t]he site ranking feature is a customized feature based on the words in the user query.” *Id.* (citing Ex. 1008 ¶ 81).

Patent Owner argues Petitioner’s assertions are insufficient because “Petitioner and its expert fail to address how Jacobsen creates this site ranking feature *in response to the customer request for information*. Instead, Petitioner attempts to rewrite the claim language to avoid addressing this limitation.” Prelim. Resp. 29–30 (citing Pet. 26–27, Ex. 1008 ¶ 65). Patent Owner further contends that “Jacobsen’s site ranking feature (the alleged second data miner) is not created in response to a keyword search query (the alleged customer request). [A person of ordinary skill in the art] would have understood that Jacobsen’s site ranking feature is a part of the ranking capability of conventional search engines, which is distinguished from data miners in the specification of the ’526 Patent.” *Id.* at 31 (citing Ex. 1008 ¶ 66). Patent Owner also asserts that “Petitioner provides no explanation of how this ‘site ranking feature’ constitutes a ‘data miner’ and a [person of ordinary skill in the art] would have understood this ‘site ranking feature’ to be different from the claimed ‘second data miner.’” *Id.* at 32–33 (citing Ex. 1008 ¶ 69). According to Patent Owner, an ordinarily skilled artisan “would understand that the miners described in the ’526 Patent do more than simply provide a keyword search and ranking feature.” *Id.* at 33–34 (*comparing* Ex. 1001, 1:46-59 (describing analysis of a question using a conventional keyword search in the background of the ’526 Patent) *with* 8:40–56 (describing analysis of the same question using data miners with a different result)).

On this record, we are persuaded by Patent Owner’s arguments. As noted above, we construe this limitation to require (1) the creation of at least one new second data miner, (2) that this creation be done in response to a customer request. *See supra* § II.B.

As to the first argument, we agree with Patent Owner that Petitioner has not shown sufficiently that Jacobsen's site ranking is created in response to a user request. Jacobsen's site ranking feature "analyzes the Internet (i.e., the sites found) to determine the degree to which sites have been found interesting by others in the desired context/profession." Ex. 1003 ¶ 67.

A word score is saved for each word on each page (i.e., except for stop words, as discussed above). When a user enters a query, the inventive SSSE determines a set of (relevant) pages that contain the query terms. For each page, the word scores are summed for the words of the query to compute a site ranking for that page. The site rankings for the pages are then used in determining how to display the search results to the user.

Id. ¶ 69. Based on our review of Jacobsen's disclosures, we are persuaded that the query is an input to the site ranking feature. As described in the '526 specification, a "'miner'" or 'data mining element' is . . . **an analysis engine** that generates an output, and specifically an output that can include one or more keys representing characteristics of an entity, using a set of rules." Ex. 1001, 5:63–66 (emphasis added). Site ranking in Jacobsen is performed after a user query, but on the record before us, Petitioner has not shown that the analysis engine used to rank the sites is not created "in response to" the user query. *See also* Pet. 14 n. 7 (stating that Jacobsen discloses "the use of a pre-existing template to search the results of a first data miner").

In addition, Petitioner has not shown sufficiently that the site ranking feature in Jacobsen would disclose a data miner as recited in the claims. The '526 patent describes prior art systems that "generat[ed] an inverted index of words that occur on Web pages." Ex. 1001, 1:22–23. User queries were provided with output based on that index and the results "returned are

usually ranked by relevance using e.g., link information or key word frequency of occurrence.” *Id.* at 1:29–31. Petitioner’s declarant describes these disclosures are applicant admitted prior art. Ex. 1008 ¶ 52. The ’526 patent criticizes this prior method as inefficient because, despite the ranking “finding particular types of information typically entails a great deal of mundane sifting through query results by a person” and this process is “labor-intensive and mundane,” ranking “can consume an excessive” amount of time.” Ex. 1001, 1:31–35, 1:42–43, 2:5–6. In contrast, the ’526 patent’s system purportedly “integrates the knowledge of many experts to sift through a large corpus of data to respond to what might be relatively complex requests for information.” *Id.* at 3:9–11. Its data miners are analysis engines that operate based on expert rules. *Id.* at code (57), 5:63–66. Petitioner’s allegations have not persuaded us that Jacobsen’s site ranking feature overcomes the problems of the prior art ranking system.

Accordingly, based on the record before us, we determine that Petitioner has not established a reasonable likelihood of prevailing on its assertion that claims 1 and 34 would have been anticipated by Jacobsen.

2. Claims 3, 19, 20, 34, 36, 52, and 53

Petitioner asserts that dependent claims 3, 19, 20, 36, 52, and 53 were been anticipated by Jacobsen. Pet. 27–32. Petitioner’s assertions do not remedy the above noted deficiency regarding the independent claims. As such, on the record before us, we are not persuaded that Petitioner’s arguments and cited evidence demonstrates a reasonable likelihood that Petitioner would prevail in its anticipation contentions based on Jacobsen with respect to claims 3, 19, 20, 36, 52, and 53.

E. Alleged Obviousness of Claims 2 and 35 Over Jacobsen, and Over Jacobsen and Jaramillo, and Alleged Obviousness of Claims 6–9, 11, 16, 17, 39–42, 44, 49, and 50 Over Jacobsen and Hegli

Petitioner asserts that dependent claims 2 and 35 are unpatentable under 35 U.S.C. § 103(a) as obvious over Jacobsen, and as obvious over Jacobsen and Jaramillo. Pet. 4, 32–38. Petitioner also asserts that dependent claims 6–9, 11, 16, 17, 39–42, 44, 49, and 50 are unpatentable under 35 U.S.C. § 103(a) as obvious over Jacobsen and Hegli. Pet. 4, 38–50. Petitioner’s assertions here rely on the same disclosures from Jacobsen regarding the above discussed independent claims and do not remedy the above noted deficiency. As such, on the record before us, we are not persuaded by Petitioner’s Jacobsen-based challenges to claims 2, 6–9, 11, 16, 17, 35, 39–42, 44, 49, and 50.

F. Alleged Anticipation of Claims 1, 6, 9, 11, 13, 19, 20, 34, 39, 42, 44, 46, 52, and 53 By Seibel

Petitioner alleges that claims 1, 6, 9, 11, 13, 19, 20, 34, 39, 42, 44, 46, 52, and 53 were anticipated by Seibel, citing the Declaration of Dr. Williams for support. Pet. 4, 50–66 (citing Ex. 1008). Patent Owner opposes, citing the Declaration of Dr. Almeroth for support. Prelim. Resp. 36–45 (citing Ex. 2001).

For reasons discussed below, we determine that Petitioner has not shown a reasonable likelihood that it would prevail in its anticipation challenge based on Jacobsen.

1. Independent Claims 1 and 34

Petitioner relies on the same disclosures from Jacobsen as to these limitations from claims 1 and 34. Pet. 62–64. Thus, this discussion applies

to Petitioner's assertions regarding both of the challenged independent claims.

Claim 1 limitations (b)–(c) recite “crawling the Internet to return Web pages to a data store” and “accessing the data store with at least a first data miner to process data in the store and output at least one result.” Ex. 1001, 11:64–67. Claim 34 similarly recites “means for accessing, with at least a first data miner, a data store containing information representing at least one Web page to process data in the store and output at least one result.” *Id.* at 13:43–46. Thus, both claims require the first data miner to access a data store that contains web pages.

Petitioner argues Seibel discloses crawling the internet with prospects harvester 21 or crawler process 31. Pet. 53. As depicted in Seibel's Figure 3, crawler 31 outputs to the “web archive,” which is an input to the “harvester.” *Id.* at 54 (reproducing Fig. 3). Prospects harvester 21 returns the information it collects from external data sources to staging database 22. *Id.* (citing Ex. 1004, 3:63–4:1). Petitioner also directs us to “database server system 41, which may be used within lead generation system 10 of Figs. 1 and 2.” *Id.* (citing Ex. 1008 ¶ 174). According to Petitioner, “Seibel discloses crawling the Internet using processes 21, 31 to return Web pages to a data store, e.g., database server system 41.” *Id.* at 55 (citing Ex. 1008 ¶ 175).

As to the first data miner accessing the data store containing web pages, Petitioner directs us to Seibel's profiles generation process 25. *Id.* Petitioner notes that “Seibel describes profiles generation process 25 which ‘analyzes the user's in-house files and records related to the user's existing customers to identify and group those customers into profile categories

based on the customer's buying patterns and purchasing volumes.” *Id.* at 56 (citing Ex. 1004, 4:23–39). Petitioner further directs us to Seibel’s disclosure of ““resulting profiles are then overlaid on the data obtained by [prospects harvester] process 21 to facilitate more accurate revenue projections and to enhance the sales and marketing process. The client may add certain value judgments (rankings) in a table that is linked to a unique lead id that can subsequently be analyzed by data mining or OLAP analytical tools.” *Id.* (citing Ex. 1004, 4:39-45).

Patent Owner argues Petitioner’s assertions do not accurately state the disclosures of Seibel. According to Patent Owner, “[t]here is no disclosure in Seibel—or explanation in the Petition—that prospects harvester process 21 returns Web pages to database server system 41. As acknowledged by Petitioner, the prospects harvester process 21 ‘harvest[s] contact information from the web and ***return[s] it to a staging data base 22.***’” Prelim. Resp. 37 (citing Pet. 53 (quoting Ex. 1004, 3:63-4:1)). Patent Owner notes that Seibel never discusses staging data base 22 outside of that one quote and states that because “there is no disclosure in Seibel that equates staging data base 22 and system 41, a [person of ordinary skill in the art] would not understand prospects harvester process 21 to return anything to database server system 41, let alone webpages.” *Id.* (citing Ex. 2001 ¶76). According to Patent Owner this is fatal to Petitioner’s challenge because “Petitioner does not show that Siebel’s profiles generation process 25 accesses staging data base 22 to meet the requirement in the claims that the alleged first data miner accesses the alleged data store to which the Web pages are returned.” *Id.*

We are persuaded by Patent Owner’s arguments and evidence. Figure 2 of Seibel is reproduced below.

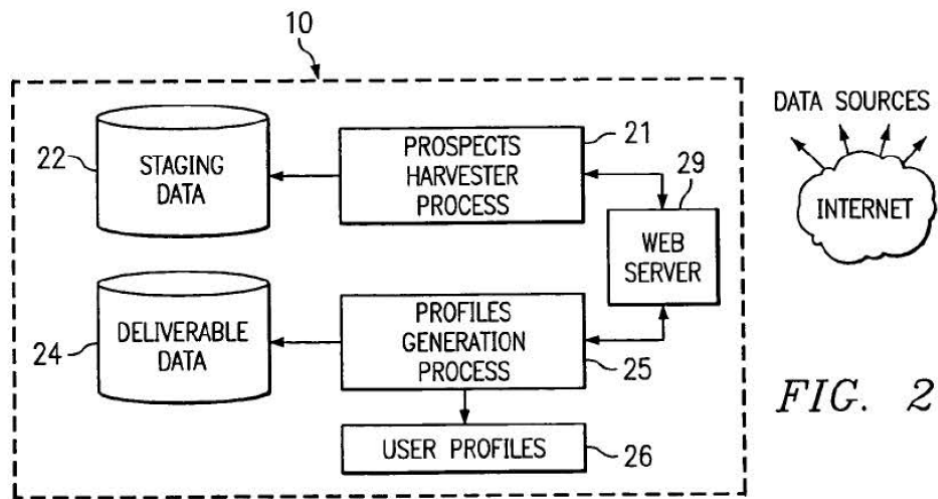


Figure 2 illustrates the functional elements of a lead generator system 10.
Ex. 1004, 2:25–27, 3:58–62.

Petitioner is relying on prospects harvester 21 to disclose the crawling of the web. As can be seen in Figure 2, prospects harvester 21 outputs to staging database 22. *See* Ex. 1004, 3:63–4:1. Petitioner is relying on profiles generation process 25 to disclose the first data miner. The only connection between prospects harvester 21 and profiles generation process 25 is web server 29. “Web server 29 provides the interface between the client systems 13 [(not shown in this Figure)] and the lead generation system 10. As explained below, it may route different types of requests to different sub processes within system 10.” *Id.* at 4:53–56. Petitioner does not direct us to nor do we see any discussion in Seibel of profiles generation process 25 accessing database 22.

Figure 3 of Seibel is reproduced below.

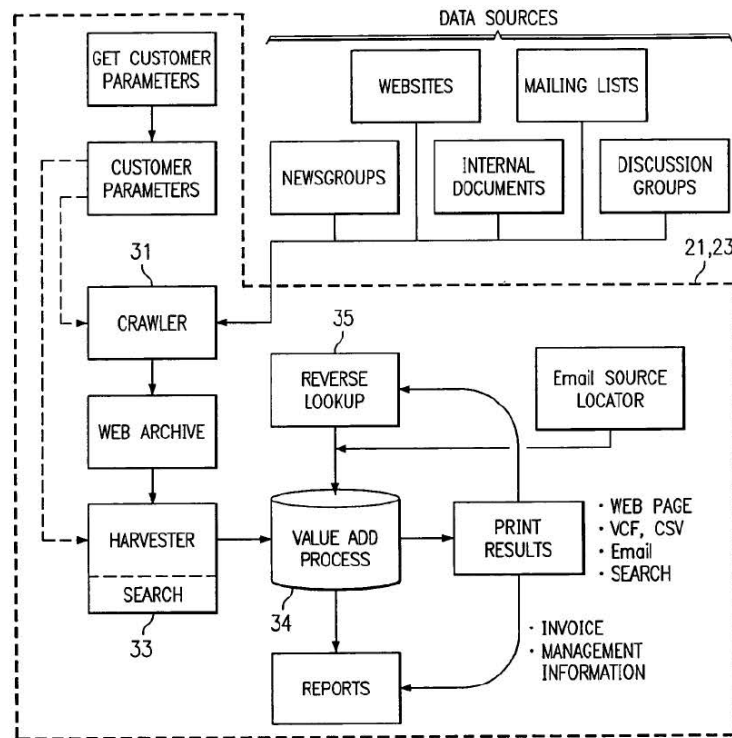


FIG. 3

Figure 3 illustrates various data sources and an embodiment of the prospects harvester. Ex. 1004, 2:28–31.

Petitioner’s reliance on web crawler 31 does not alleviate the deficiency discussed above. Petitioner does not direct us to any discussion of where the output from web crawler 31 is stored. Based on our review of Seibel, the only output that we see is to the “web archive,” which feeds into the “harvester.” *See id.* at Fig. 3. The unnumbered element “web archive,” is not specifically discussed in the text and is shown as a box in between web crawler 31 and harvester/search 33. Element 33 of Figure 3, “Harvester – Search,” is not specifically referenced in Seibel’s text. A number of elements of Figure 3 are located within a dotted line box labeled “21, 23,” however our review of Seibel reveals no discussion of element “23.” Thus, based on our review of Seibel, web crawler 31 may be an element of

prospects harvester 21 or an input to the prospects harvester. In either event, we see no disclosure in Seibel of web crawler 31 storing web pages into a data store accessible by profiles generation process 25.

Finally, Petitioner's reliance on database server system 41 does not remedy this deficiency. Seibel states that database server system 41 "could be used to implement the profiles generation process 25, which collects profile data about the client." *Id.* at 5:62–64. Database server system 41 is composed of data acquisition component 41a, data mart 41b, predefined query repository 41c, business intelligence charts repository component 41d, and web server component 41e. *Id.* at 6:11–7:28. Petitioner specifically references data acquisition component 41a noting that it "collects and extracts data from different data sources." Pet. 54 (citing Ex. 1004, 6:12–18). The input for data acquisition component 41a, however, is depicted as input data 42. *See* Ex. 1004, Fig. 4. Input data 42, however, is not described as containing web pages. As described in Seibel,

The input data 42 can be the client's sales data, customer-contact data, customer purchase data and account data etc. Various data sources for customer data can be contact management software packages such as ACT, MarketForce, Goldmine, and Remedy. Various data sources for accounting data are Great Plains, Solomon and other accounting packages typically found in small and medium-sized businesses. If the client has ERP (enterprise resource planning) systems (such as JD Edwards, PeopleSoft and SAP) installed, the data sources for customer and accounting data will be extracted from ERP customer and accounting modules. This data is typically structured and stored in flat files or relational databases.

Id. at 5:65–6:10; *see also* Ex. 2001 ¶¶ 47, 80 (opining that system 41 includes input from internal data sources). This comports with the description of profiles generation process 25 which states that it "analyzes

the user's in-house files and records related to the user's existing customers to identify and group those customers into profile categories based on the customer's buying patterns and purchasing volumes.” *Id.* at 4:23–27 (emphasis added). Thus, based on the record before us, we are not persuaded by Petitioner's allegations that Seibel discloses a first data miner that accesses a data store that contains web pages.

Independent claim 1 and 34 both recite “creating at least a second data miner in response to the customer request for information, the second data miner accessing at least the result of the first data miner to generate an output.” Ex. 1001, 12:1–4 (claim 1), 13:48–51 (claim 34). Petitioner argues Seibel discloses the claimed “creating.” Pet. 57–59, 64.

As to the recited second data miner, Petitioner directs us to Seibel's server 41e, which “runs a query using different parameters (e.g., a quarterly revenue query where the parameters passed are the specific region, the specific year, and the specific quarter) to extract data from data mart 41b to feed the data to the requested chart.” Pet. 57. Petitioner asserts that server 41e performs “a customized extraction based on the parameters of the user request.” *Id.* Patent Owner argues that “Petitioner and its expert fail to address how Seibel creates this server *in response to the customer request for information.*” Prelim. Resp. 41 (citing Pet. 57–59; Ex. 2001 ¶ 83).

On this record, we are persuaded by Patent Owner's arguments. As noted above, we construe this limitation to require (1) the creation of at least one new second data miner, (2) that this creation be done in response to a customer request. *See supra* § II.B. As described in Seibel, Web Server component 41e is a component of system 41 and the components of system 41 are described as commercially available servers such as Microsoft IIS,

Apache, Tango, Cold Fusion or PHP. Ex. 1004, 7:11–21. Petitioner does not argue that the customer request led to the creation of server 41e, but rather that the parameters passed to 41e are customized based on the customer request. Pet. 57; *see also id.* at 14 n. 7 (stating that Seibel discloses “the use of a pre-existing template to search the results of a first data miner”). This is not sufficient to disclose the recited creation of a second data miner. As discussed above, the intrinsic record persuades us that creation of a second data miner requires the creation of a new “analysis engine” and not just the provision of customized parameters to an existing “analysis engine.” *See supra* § II.B (citing Ex. 1001, 5:63–66, 6:25–26, 6:36–40). Thus, Petitioner’s allegations and evidence are insufficient to show that Seibel discloses the recited creation of a second data miner.

Accordingly, based on the record before us, we determine that Petitioner has not established a reasonable likelihood of prevailing on its assertion that claims 1 and 34 would have been anticipated by Seibel.

2. Claims 6, 9, 11, 13, 19, 20, 34, 39, 42, 44, 46, 52, and 53

Petitioner asserts that claims 6, 9, 11, 13, 19, 20, 39, 42, 44, 46, 52, and 53 would have been anticipated by Seibel. Pet. 60–66. Petitioner’s assertions do not remedy the above noted deficiency regarding the independent claims. As such, on the record before us, we are not persuaded that Petitioner’s cited evidence demonstrates a reasonable likelihood that Petitioner would prevail in its anticipation contentions based on claims 6, 9, 11, 13, 19, 20, 39, 42, 44, 46, 52, and 53 as anticipated by Seibel.

3. Alleged Obviousness of Claims 2 and 35 Over Seibel

Petitioner asserts that claims 2 and 35 are unpatentable under 35 U.S.C. § 103(a) as obvious over Seibel. Pet. 4, 66–68. Petitioner’s

assertions here rely on the same disclosures from Seibel regarding the above discussed independent claims and do not remedy the above noted deficiency. As such, on the record before us, we are not persuaded by Petitioner's allegations and evidence that claims 2 and 35 would have been obvious over Seibel.

III. DISCRETIONARY DENIAL

Patent Owner argues that we should exercise our discretion under 35 U.S.C. § 314(a) and not institute trial, in view of parallel district court litigation in the Eastern District of Texas. Prelim. Resp. 16–28. We do not reach those issues in light of our assessment of the merits of the Petition, as discussed above.

IV. CONCLUSION

For the foregoing reasons, we have determined that Petitioner has not established a reasonable likelihood of prevailing with respect to any claims. Thus, we do not institute an *inter partes* review.

V. ORDER

In consideration of the foregoing, it is hereby:

ORDERED that, pursuant to 35 U.S.C. § 314(a), the Petition is denied, and no *inter partes* review is instituted.

IPR2023-01441
Patent 7,010,526 B2

For PETITIONER:

Thomas M. Dunham
Robert W. Molitors
Elizabeth A. O'Brein
CHERIAN LLP
tomd@cherianllp.com
rmolitors@cherianllp.com
elizabetho@cherianllp.com

For PATENT OWNER:

Eagle H. Robinson
Daniel S. Leventhal
Chad Wallis
NORTON ROSE FULBRIGHT US LLP
eagle.robinson@nortonrosefulbright.com
daniel.leventhal@nortonrosefulbright.com
chad.wallis@nortonrosefulbright.com