

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

BRIGHT DATA LTD.,

Petitioner

v.

OXYLABS, UAB,

Patent Owner

Case IPR2023-01442
Patent No 7,020,667

PATENT OWNER PRELIMINARY RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	OVERVIEW OF THE '667 PATENT.....	1
A.	Operational Architecture and Navigation.....	2
1.	<i>Initialization Stage</i>	4
2.	<i>Automatic Execution Stage</i>	5
B.	Independent Claims 1, 16, and 20.....	6
III.	OVERVIEW OF THE PRIOR ART.....	8
A.	DaCosta (EX1004)	8
1.	<i>Navigation Module</i>	10
2.	<i>Extraction Module</i>	12
3.	<i>Smart Server</i>	14
B.	Kraft (EX1005)	16
IV.	RELATED LITIGATION	17
V.	INSTITUTION SHOULD BE DENIED PER <i>FINTIV</i>	18
A.	Petitioner Has Not Requested a Stay Based on Its IPR Petition.....	20
B.	The District Court Trial Will Occur Months Before the Board Could Reach a Final Written Decision in this Proceeding.....	21
C.	The District Court and Parties Have Invested and Will Invest Significantly in the Litigation.....	22
D.	The Petition and the Parallel Litigation Will Have Significant Overlap That is Not Mitigated by Petitioner's Stipulation.....	23
1.	<i>Prior Art & Invalidity Theories</i>	23
2.	<i>Additional Claims</i>	25
E.	The Parties Are the Same in the Litigation.....	26
F.	The Petition Lacks Compelling Merits.....	26

VI.	GROUND 1: THE PETITION FAILS TO SHOW ANTICIPATION OF CLAIMS 1, 16, AND 20 BY DACOSTA.....	29
A.	Claims 1, 16, and 20 are Novel Over DaCosta.....	30
1.	<i>[1g], [16g], [20g]: selecting the target database and target locations in the target database where the at least one of the data elements is to be stored in the target database.....</i>	<i>31</i>
2.	<i>[1i], [16i], [20i] – “setting a schedule frequency for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database”.....</i>	<i>40</i>
3.	<i>[1h], [16h], [20h] “saving the source locations and target locations of the at least one of the parsed data elements in a location file in the target database;”</i>	<i>43</i>
4.	<i>[16b], [16j]: First and second sets of instruction codes.....</i>	<i>48</i>
VII.	GROUND 2: THE PETITION FAILS TO SHOW ANTICIPATION OF CLAIMS 1, 16, AND 20 BY KRAFT	49
A.	Claims 1, 16, and 20 are Novel Over Kraft.....	49
1.	<i>[16b], [16j]: First and second sets of instruction codes</i>	<i>50</i>
2.	<i>[1i], [16i], [20i]: – “setting a schedule frequency for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database;”.....</i>	<i>50</i>
3.	<i>[1e], [16e], [20e]: – “selecting at least one of the data elements from the list to be stored in the target database”</i>	<i>53</i>
4.	<i>[1d], [16d], [20d]: “parsing selected data elements in the source file into a list of data elements;”.....</i>	<i>54</i>
VIII.	CONCLUSION.....	58

PATENT OWNER’S EXHIBIT LIST

Exhibit	Shorthand	Description
2001	Almeroth	Declaration of Kevin Almeroth, Ph.D.
2002	CV	CV of Kevin Almeroth, Ph.D.
2003	Gilstrap	List of Decisions for Motions to Stay Pending Inter Partes Review by Judge Gilstrap since Jan. 1 st 2023
2004	Docket	Civil Docket for Case EDTX-2-23-cv-00171
2005	Linux	James Mohr, <i>Introduction to Operating Systems – Processes</i> , The Linux Knowledge Base and Tutorial (2004) (archived 2007) (available at https://web.archive.org/web/20071011225814/http://www.linux-tutorial.info/modules.php?name=MContent&pageid=3)
2006	NinjaOne	Lauren Ballejos, <i>What Is a Background Process?</i> , NinjaOne (updated 2023) (available at https://www.ninjaone.com/blog/what-is-a-background-process/#:~:text=A%20background%20process%2C%20as%20the,require%20user%20intervention%20to%20function.)
2007	Merriam-Webster	“Preset.” Merriam-Webster.com Dictionary, Merriam-Webster, (https://www.merriam-webster.com/dictionary/preset). Accessed 12 Jan. 2024.
2008	Sundaresan	U.S. Patent No. 6,606,620 to Sundaresan
2009	Ehrlich	U.S. Patent No. 6,873,968 to Ehrlich
2010	Miner	U.S. Patent No. 7,653,702 to Miner
2011	Dill	U.S. Patent No. 7,730,013 to Dill

I. INTRODUCTION

Institution should be denied because the Petition fails to establish a reasonable likelihood of prevailing on any challenged claim. The references cited in each of the Grounds fail to disclose each of the features as recited in the claims; the Petition improperly attempts to plug the gaps in its anticipation theories by recasting distinct components in the references as purportedly interchangeable (despite any disclosure of such interchangeability in the references); and Dr. Williams' analysis is conclusory and legally insufficient to supply the deficiencies.

Further, to the extent the Board finds that any of Petitioner's grounds chins the bar of a reasonable likelihood, the Board should still exercise its discretion to deny institution in view of the co-pending litigation. Moreover, the Petition's overall lack of analysis and Petitioner's failure to show how a POSITA would understand the features of the prior art as corresponding to the claimed features are insufficient to meet the compelling merits standard.

For each of these reasons, institution should be denied.

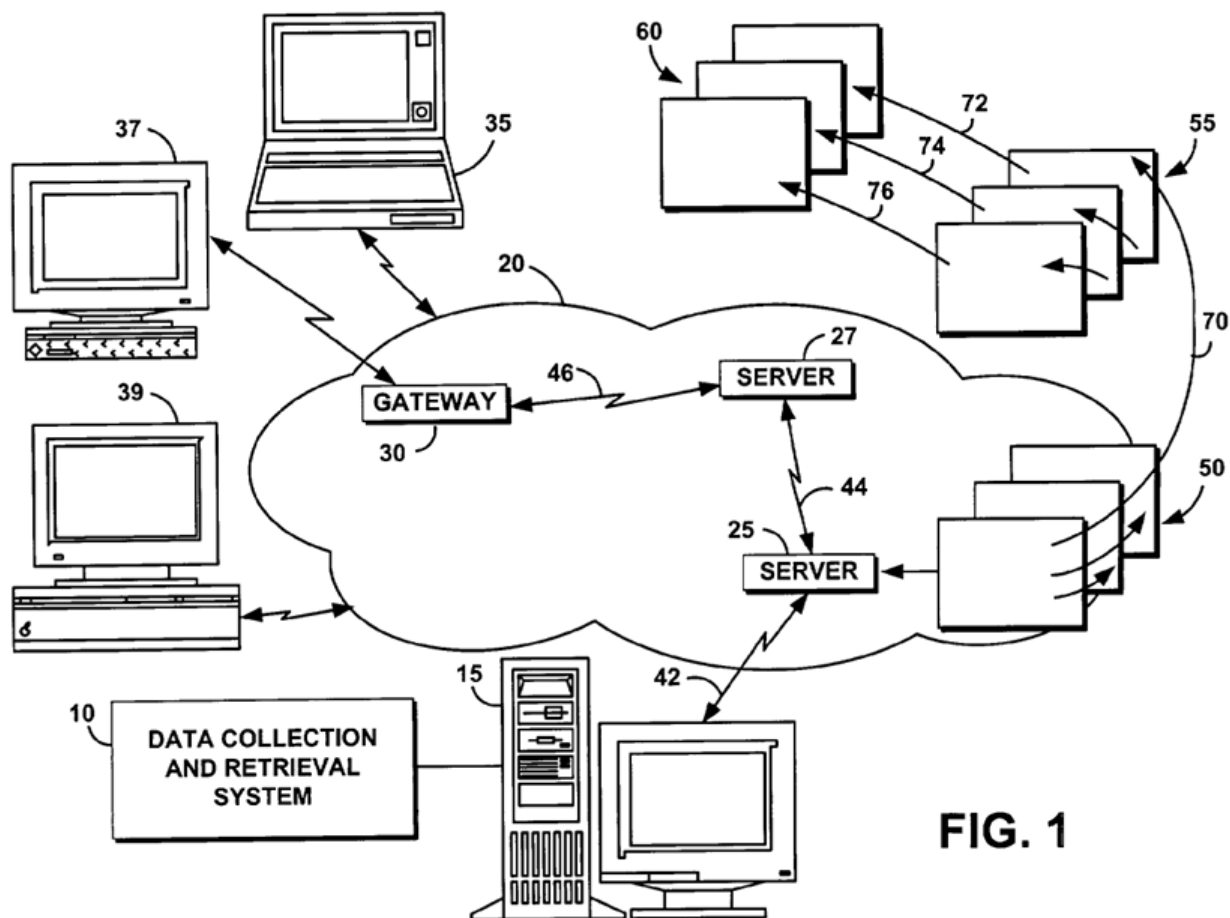
II. OVERVIEW OF THE '667 PATENT

The '667 Patent describes a "data collection and retrieval system and associated method [to] allow the capture and replication of data presented at various Web pages into a database application." EX1001, Abstract. The system can include (1) an initialization stage—which provides the user interface that allows the user to

select the source file that contains data the user wishes to copy, a target database that will receive the data, and timing criteria for automatic transfer of data—and (2) an automatic execution stage that automatically transfers the data from the source file to the target database as instructed by the user in the initialization stage. *Id.*

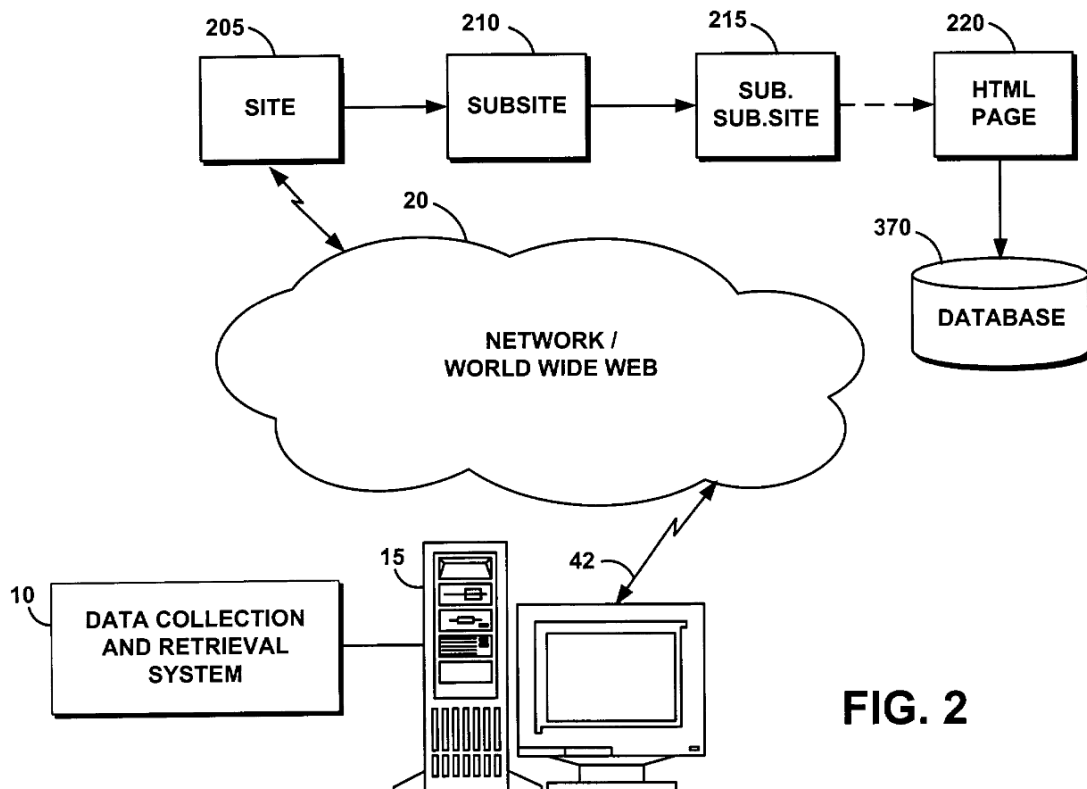
A. Operational Architecture and Navigation

An illustrative operating environment in which the data collection and retrieval system can be used is depicted in the component diagram of FIG. 1 (inset below). EX1001, 2:50-53.



As shown in FIG. 1, “servers 25, 27 provide access to stored information such as hypertext or Web documents indicated generally at 50, 55, and 60” that can include hypertext links (e.g., 70, 72, 74, 76) to other web sites or documents that are stored by various Web servers. EX1001, 4:6-16.

In this way, the user can navigate to various webpages “until a webpage ... is found that contains the data in which the user is interested.” *Id.* 4:17-26. An example of this navigation path is shown in FIG. 2 (inset below), in which a user selects different subsites 210, 215 before arriving at particular webpage 220 that the user wishes to transcribe (or copy) to a database. *Id.*



Once a user has navigated to (or otherwise identified) the desired webpage (220), the user can implement the initialization stage 305 and automatic execution stage.

1. *Initialization Stage*

An illustrative method of performing the initialization stage of the system is shown in the flow chart found in FIG. 4 (inset below) of the '667 Patent.

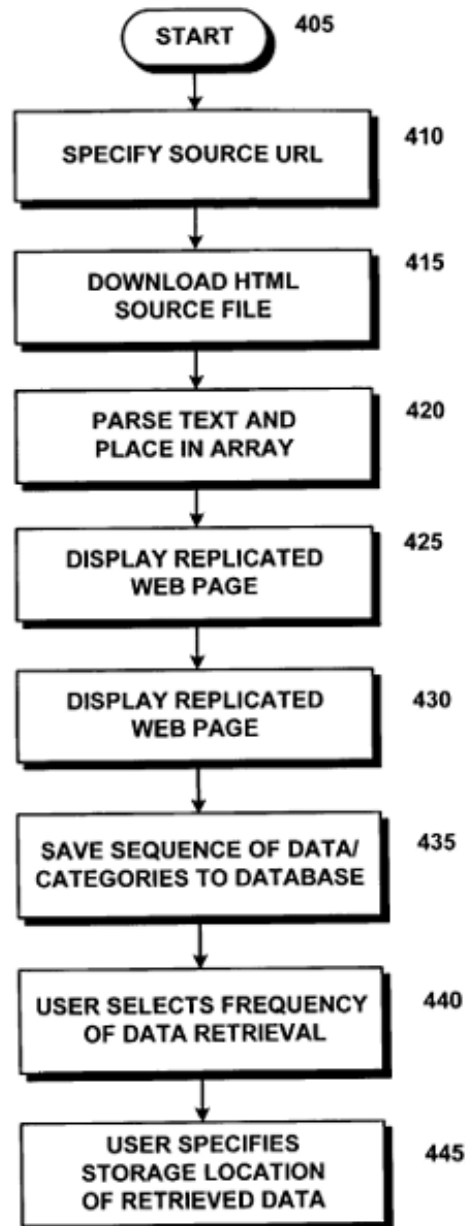


FIG. 4

EX1001, 2:61-62. This illustrative method starts with step 410, where “the user specifies the URL for an HTML page 220 that contains data to be transcribed to the

target database 370.” *Id.*, 4:46-49. This step corresponds to the navigation to web page 220 described above in Section II.A. *Id.*, 4:46-49 (“user selects the Web page 220, as shown by FIG. 2, and saves the Web page URL in system 10.”).

After the user has found the desired webpage (e.g., 220), the system will then “download the HTML source file 320 for the Web page 220 in step 415” and a parser 415 will process the source file 320, tokenize the HTML format of the Web page 220, and place it in a data array 330 in step 420. EX1001, 4:56-59. The system may then replicate and display portions of the webpage in a refined user GUI 340. *Id.*, 4:59-66, 5:6-11. The user can then select data and or categories of interest (e.g., from the replicated webpage) in step 430 and the system can save the location of the selected data so that it may retrieve corresponding data from the same location during successive data transfer. *Id.*, 5:12-16. Finally, the user can select “the frequency of data retrieval” at step 440 and specify where to store the retrieved data. *Id.*, 5:17-29.

2. Automatic Execution Stage

After the initialization stage 305 is complete, the system can then perform the automatic retrieval of data. EX1001, 5:30-35. An illustrative example of operation of the automatic execution stage is depicted in FIGs. 5A-5B. *Id.* For example, the data retrieval process can begin after expiration of the time specified by the user in the initialization stage by accessing the web page 220. *Id.*, 5:35-37. The system

may then parse the source file of the identified webpage, extract data from the webpage that corresponds to the data selected by the user, and transfer that data to a database (e.g., 370) that was previously specified by the user. EX1001, 5:37-45. This process is then repeated according to the frequency that was previously specified by the user. *Id.*, 5:45-47.

B. Independent Claims 1, 16, and 20

Independent claim 1 of the '667 Patent is directed to a “method of selectively retrieving data from a source file ... and automatically collecting retrieved data in a target database.” EX1001, 6:63-65. FIGs. 4 and 5A of the '667 Patent, described above, show partial flowchart diagrams of an illustrative method of retrieving and collecting data for some of the steps covered by independent claim 1.

Claim 1 recites:

- [1a] A computer-implemented method of selectively retrieving data from a source file on a public network and automatically collecting retrieved data in a target database, comprising:
- [1b] initializing a data retrieval stage by:
- [1c] selecting the source file;
- [1d] parsing selected data elements in the source file into a list of data elements;
- [1e] selecting at least one of the data elements from the list to be stored in the target database;
- [1f] identifying source locations of the selected at least one of the data elements;

- [1g] selecting the target database and target locations in the target database where the at least one of the data elements is to be stored in the target database;
- [1h] saving the source locations and target locations of the at least one of the parsed data elements in a location file in the target database;
- [1i] setting a schedule frequency for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database; and
- [1j] automatically, selectively parsing new data elements corresponding to the at least one of the parsed data elements previously selected during the initialization of the data retrieval stage, pursuant to the schedule frequency, by transferring the selectively parsed new data elements from the saved source locations in the source file to the saved target locations in the target database;

EX1001, 6:63-7:25.

Claim 16 includes limitations similar to the Limitations [1b]-[1j] of claim 1, and recites **[16a]** a computer program product having instruction codes for selectively retrieving data from a source file on a public network and for automatically collecting retrieved data in a target database, comprising: **[16b]** “a first set of instruction codes for initializing a data retrieval stage by:” performing Limitations [1a]-[1i] and **[16j]** “second set of instruction codes for” performing Limitation [1j]. *Id.*, 7:61-8:26.

Claim 20 also limitations similar to the Limitations [1b]-[1j] of claim 1 and recites **[20a]** a computer-implemented system for selectively retrieving data from a source file on a public network and for automatically collecting retrieved data in a target database, comprising **[20b]** means for initializing a data retrieval stage and **[20j]** means for automatically, selectively parsing new data elements. *Id.*, 8:34-64.

III. OVERVIEW OF THE PRIOR ART

A. DaCosta (EX1004)

DaCosta discloses a system for “automatically extracting data from at least one electronic document accessible through the Internet or other computer network.” EX1004, Abstract. Using DaCosta’s system, “one can automatically navigate to a plurality of web site destinations, extract specified information based upon taught schemas, process the extracted data according to customizable scripts, integrate information from other applications such as Microsoft (“MS”) Word, Excel, or Access, [and] view the final output using a browser such as Microsoft Internet Explorer.” *Id.*, 5:12-19. These steps can then be automatically repeated. *Id.*, 5:19-20.

FIG. 1 of DaCosta (inset below) depicts an embodiment that includes a navigation API 10 (i.e., navigation module 10) and an extraction API 20 (i.e., extraction module 20). *Id.*, 5:37-67. “[T]he navigation API 10 provides two services: recording of navigation paths 12 and playback of that which has been

recorded 14.” *Id.*, 6:50-53. Additionally, the extraction API 20 “provides two services: recording of extraction patterns 22 and playback of extraction patterns 24.” *Id.*, 6:53-55.

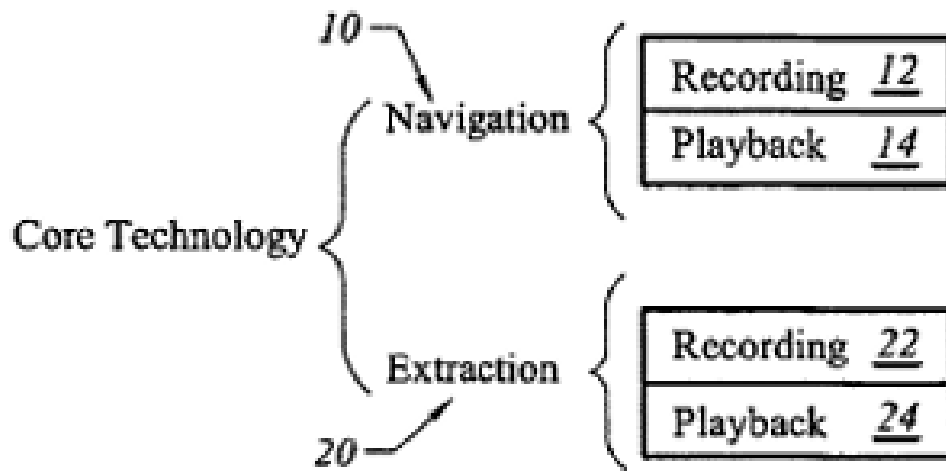


FIG. 1

The navigation module 10 is distinct from the extraction module 20, and each perform different processes. Almeroth, ¶49; EX1004, 8:67-9:4 (The extraction module 20 “follows the navigation module 10”). “[W]hile the navigation module [10] operates across multiple web pages or web space, the extraction module 20 operates at the HTML or web page or other web-accessible document level.” EX1004, 8:61-65; *see also id.*, 5:61-63.

The distinct navigation module 10 and extraction module 20 can be implemented via separate “plug-ins in a browser.” Almeroth, ¶50; EX1004,

10:18-20. Additionally, FIG. 7 shows an example of an application 220 that could implement these two separate modules (10, 20).

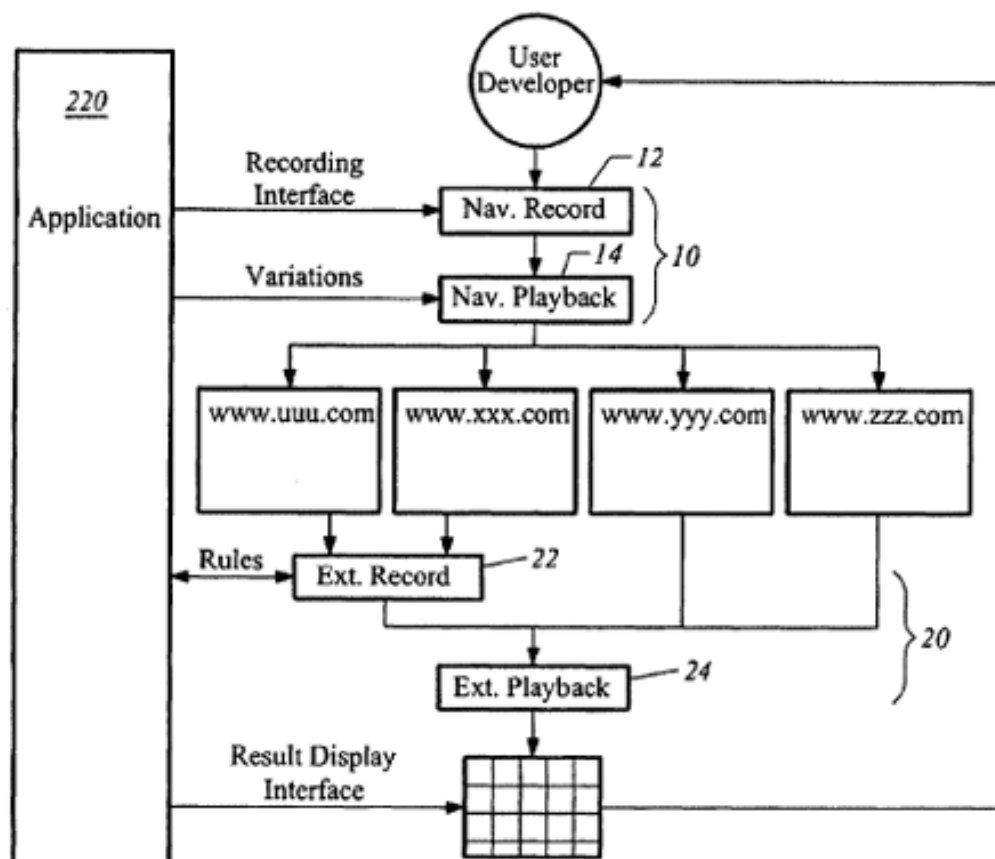


FIG. 7

EX1004, 10:55-57.

1. Navigation Module

Navigation module 10 enables “a user to learn and store navigation paths to given web pages.” EX1004, 5:37-40. “For example, if a web site requires a user to enter a login name and password to reach an orientation page and then asks for a set of preferences to go to specific web pages or other web-accessible documents of

interest, it is an object of the present invention to enable a client application to record this path once, then play it back many times.” *Id.*, 5:44-50.

FIG. 2 (inset below) depicts a process of recording navigation paths with the navigation recording module 12. *Id.*, 6:57-59.

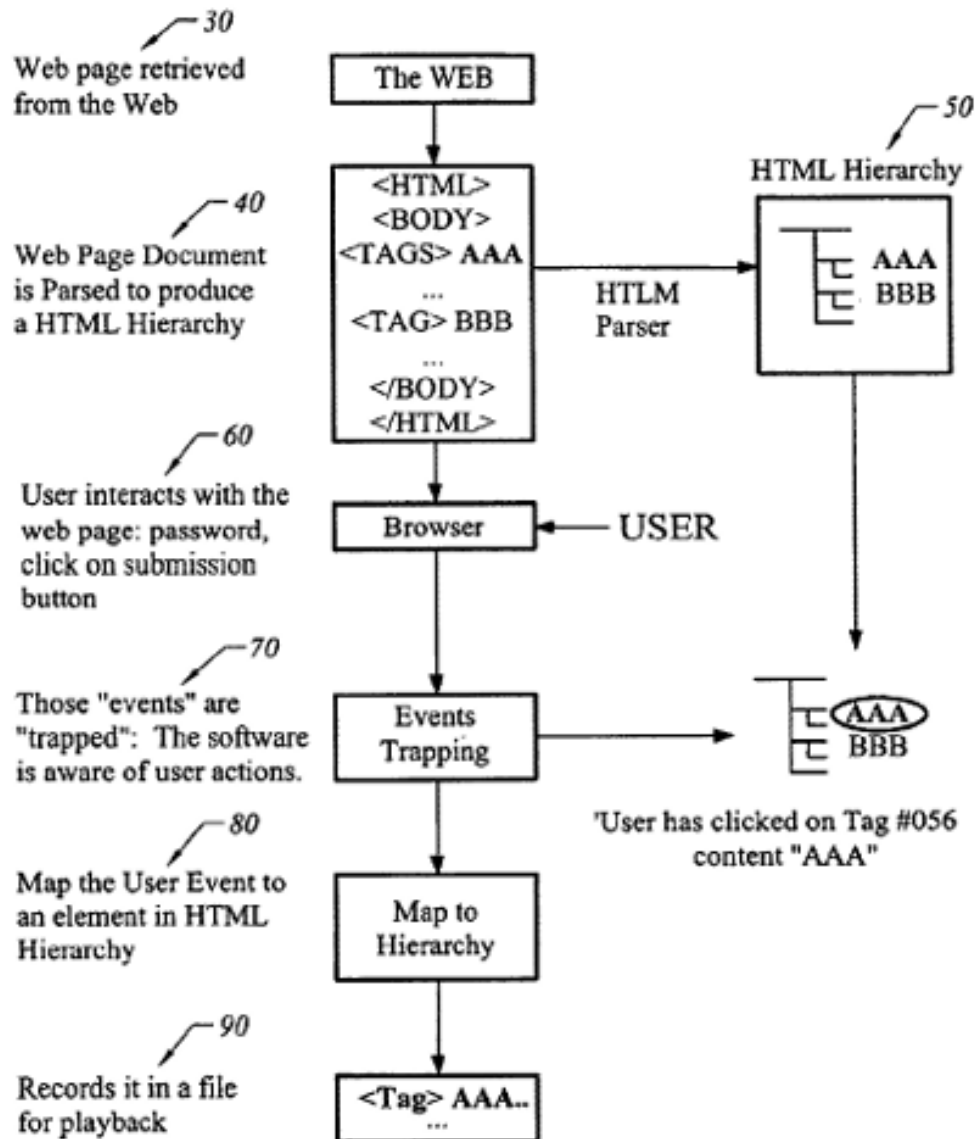


FIG. 2

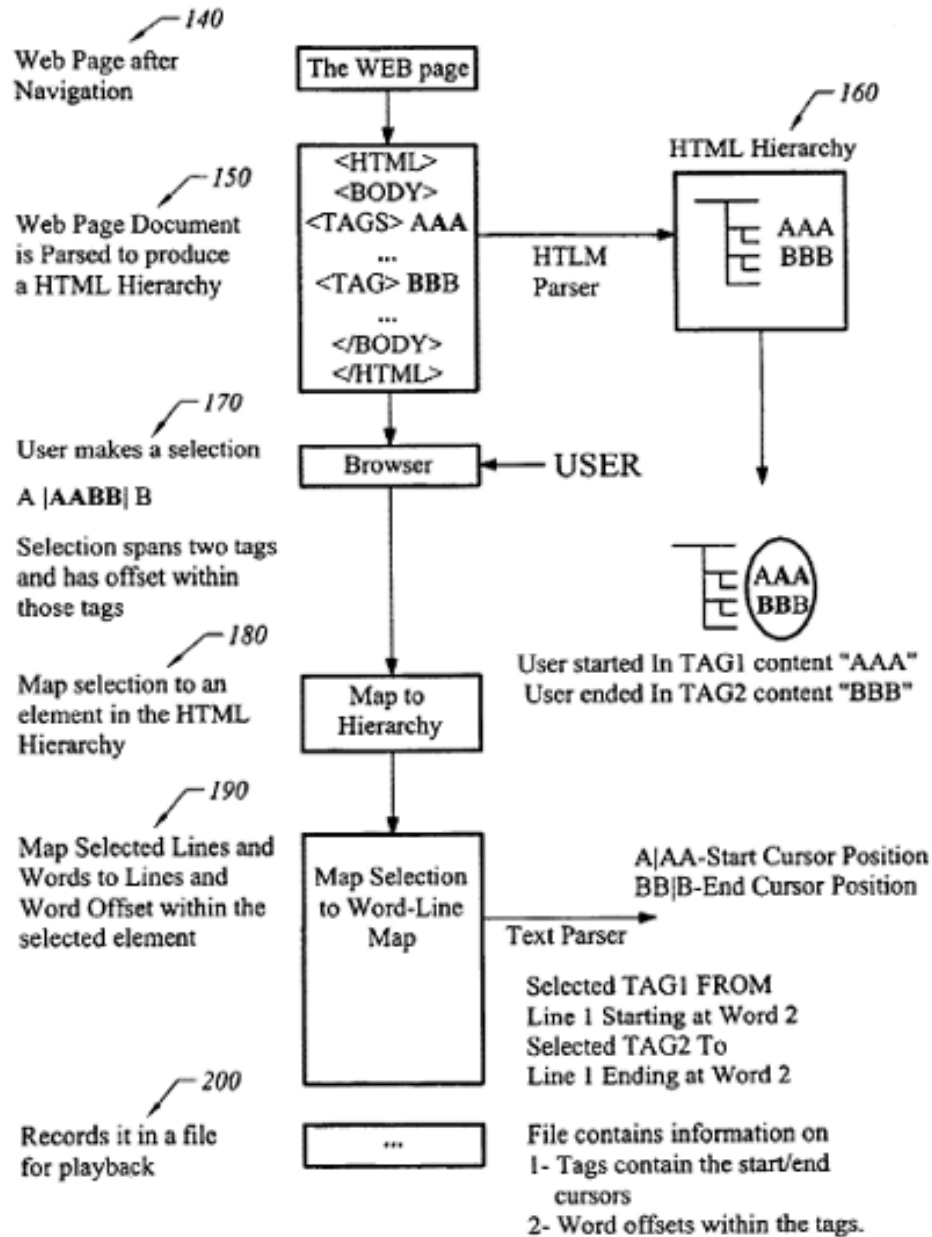
In this process, a web page is provided to navigation module (e.g., via a web browser) and the web page is parsed to convert the HTML file into an object hierarchy. EX1004, 6:60-7:9. The user then interacts with the web page by clicking on links or buttons or otherwise entering data, which is recorded and saved in an Extensible Markup Language (“XML”) file. *Id.*, 7:17-34.

The recorded navigation XML file can then be used as an input file to the navigation playback module 14. *Id.*, 8:19-22. This “navigation XML file contains details about the recorded navigation in a format that can be read by the navigation playback module 14” and can include a series of steps that the user took to reach the page. *Id.*, 7:41-51 (steps include information such as hyperlinks and form fields that are acted upon). Playback module 14 performs “two different types of steps”—(1) form variations and (2) URL variations—which allow a user to change how the navigation script is played back. *Id.*, 7:55-57. For example, form variations allow form parameters can be changed by “specifying new form parameters to replace the ones originally recorded using the navigation recording module 12.” *Id.*, 7:57-63; Almeroth, ¶54 (form variations may include updating or changing information submitted to a form field, such as updating a password or changing queries).

2. Extraction Module

Extraction module 20 “follows the navigation module 10 by processing each final or target web page ... to extract relevant or select information or data therefrom”

EX1004, 8:67-9:4. DaCosta's FIG. 4 (inset below) depicts a process for recording an extraction script via extraction recording module 22.



Recording Extraction Script - What to get from the page

FIG. 4

In this process, a web page is parsed to generate the HTML hierarchy.

EX1004, 9:53-55. A user may then select text from the web page, such as by

highlighting a portion of that web page. *Id.*, 9:57-59. This selection is then mapped, and the script is recorded as an XML file that will be input to the extraction playback module 24. *Id.*, 10:1-8. Extraction playback module 24 can then extract actual information from the web page as a text range 210. *Id.*, 9:57-59.

This extracted data is then output to the browser of the user. Almeroth, ¶57; EX1004, 5:12-20 (the user “view[s] the final output using a browser such as Microsoft Internet Explorer”); *see also* EX1004, FIG. 7 (depicting extraction playback module 24 sending outputting data to the “result display interface”), 10:45-47 (A calling application arranges the output “in a form suitable for display to the user”), FIG. 10 (“view output on your browser”).

3. *Smart Server*

In addition to the above-described modules, DaCosta discloses that its system may also include a smart server to gather data on the Internet. EX1004, 18:8-10; Almeroth, ¶58. The role of the “smart server is to fill a knowledge repository in *background mode*.” EX1004, 18:8-12.¹ DaCosta does not disclose the smart server storing data in the knowledge repository other than in this “background mode.” Almeroth, ¶58. Unlike the navigation module 10 and extraction module 20—that require a user to interact with the page and select the text desired to be extracted (EX1004, 7:16-28, 9:57-59)—the smart server operates in the background *without*

¹ All emphasis added unless otherwise noted

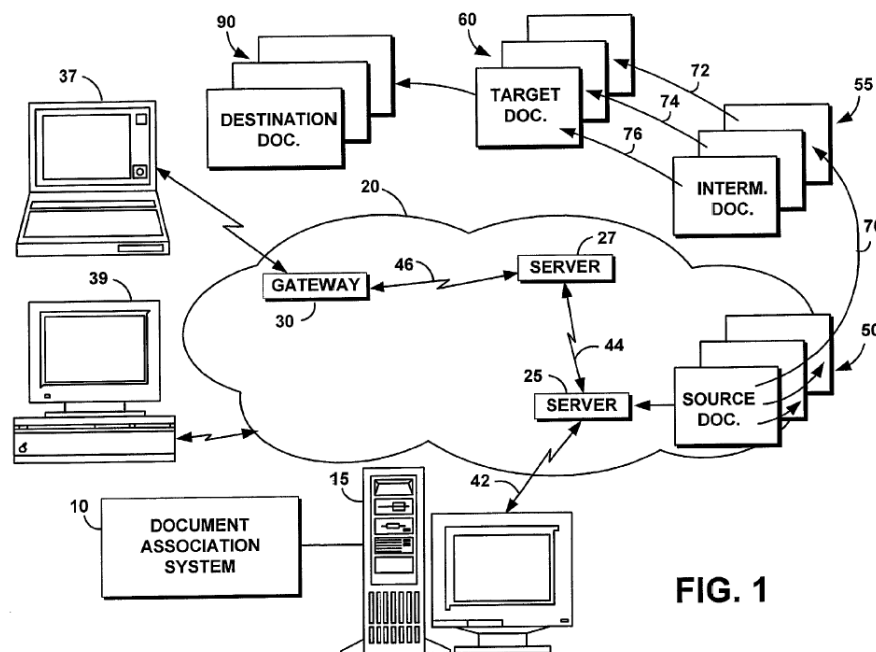
the active interaction of a user. Almeroth, ¶58. After data is stored in the knowledge repository during this background operation, a user can then access the data stored within the knowledge repository. EX1004, 18:15-23 (the user can readily access the stored data). Thus, recording and playback of user generated navigation and extraction scripts is not related to the operation of the smart server. Almeroth, ¶58.

A POSITA would understand this smart server is distinct from the user-oriented processes described in DaCosta's FIG. 1-19. Almeroth, ¶59; EX1004, 15:55-66 ("The description [of FIGs. 1-19] refers to various embodiments developed to aid in data extraction from document sources, typically (though not exclusively) accessed on the Internet through a browser."). The smart server that is depicted and described with respect to FIG. 22 does not employ the same processes disclosed in these earlier figures. Almeroth, ¶60. For example, FIGs. 1-15 and 21 describe client oriented features of the "preferred embodiment of the present invention," while FIG. 22 "illustrates an architecture for a smart web server to scour data from the Internet for deposit in a data repository." See EX1004, 4:10-5:3. DaCosta intentionally changes the language in the description of FIG. 22 to not take antecedent basis from the "preferred embodiment" described in the previous figures. EX1004, 4:10-5:3. A POSITA would understand this change in terminology and functionality as teaching that the smart server is distinct from recording of the scripts shown in FIGs. 2 and 4 and the system shown in FIG. 7. Almeroth, ¶60;

B. Kraft (EX1005)

Kraft is an application filed by the same Applicant as the '667 Patent. EX1001, 1/13 (Assigned to “International Business Machines Corporation”); EX1005, 1/13 (same). Accordingly, Kraft and the '667 Patent share some of the same background disclosure, such as the definitions found in paragraphs [0025]-[0039] of Kraft. *Compare* EX1001, 3:4-53 *with* EX1005, ¶¶[0025]-[0039].

Kraft discloses “a system for saving the content of a target document bundled with contextual metadata, such as the location of a source document, as attributes of the target document.” EX1005, ¶[0001]. Referring to FIG. 1 (inset below), Kraft depicts the system 10 in an exemplary computing environment in which servers 25, 27 provide access to stored information such as hypertext or web documents indicated generally at 50, 55, and 60.



EX1005, ¶¶[0040]-[0042]. A user can then navigate between websites or documents until the user “detects the desired target document 60.” *Id.*, ¶¶[0057]-[0059]. After navigating to the target document, the user “issues a save command, and enters the destination address (URL) 176 of the destination documents repository 170 where he or she desires to store the destination document 90.” *Id.* The destination document corresponds to the entire target document 60 along with additional “context data,” which includes the URL 630 of the source document 50, the URL 640 of the target document, the author, and a revision number. *Id.*, ¶[0060]. Thus, Kraft’s system can save an entire webpage (e.g., target document 60) along with information as to how a user navigated to that webpage (e.g., context data) in a document repository. Almeroth, ¶63.

IV. RELATED LITIGATION

Patent Owner Oxylabs, UAB, filed suit against Petitioner, Bright Data Ltd., on January 07, 2022, in the referenced action (2:22-cv-00011-JRG-RSP). *See e.g.*, Pet., 71 n.2. Bright Data asserted a Counterclaim on April 14, 2023, that was severed into a new action (2:23-cv-00171-JRG-RSP) that involves United States Patent No. 7,020,667 (the “667 patent”). *Id.*

In this current action, both initial invalidity contentions and amended invalidity contentions have already been served (EX2004, 2-3), and the exchange of preliminary claim constructions are scheduled to be completed by March 12, 2024,

before the Board is expected to issue its institution decision by April 17th, 2024. EX1007, 4/7. Trial is also expected to occur in late 2024. Pet., 71 (“trial would statistically occur around September 2024”) (citing EX1008 at 35); *see also* EX1007, 1/7 (trial set for December 9, 2024). Thus, trial is expected to begin 4-7 months before the Board is expected to issue any final determination in this matter, around April 2025.

V. INSTITUTION SHOULD BE DENIED PER *FINTIV*

Discretionary denial under Section 314(a) is appropriate here because the institution would be a highly inefficient use of the Board’s resources in view of the pending Litigation, and the Petition does not present sufficient grounds—much less *compelling evidence*. Section 314(a) vests the Director with discretion to deny institution. *Cuozzo Speed Techs., LLC v. Lee*, 136 S. Ct. 2131, 2140 (2016). As explained in Section IV above, the parties are already litigating infringement and validity of the ’526 Patent in district court litigation that will reach trial months before this proceeding could reach a final written decision.

The Board considers the presence and status of parallel district court litigation. *NHK Spring Co., Ltd. v. Intri-plex Techs.*, IPR2018-00752, Paper 8 (PTAB Sept. 12, 2018) (precedential) (“NHK”) and *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 at 5-6 (PTAB Mar. 20, 2020) (precedential, designated May 5, 2020) (“*Fintiv I*”);

Apple Inc. v. Fintiv, Inc., IPR2020-00019, Paper 15 at 7-8 (PTAB May 13, 2020) (“*Fintiv II*”).

Fintiv sets forth factors balancing system efficiency and fairness for parallel proceeding(s):

1. whether the court granted a stay or evidence exists that one may be granted if a proceeding is instituted;
2. proximity of the court’s trial date to the Board’s projected statutory deadline for a final written decision;
3. investment in the parallel proceeding by the court and the parties;
4. overlap between issues raised in the petition and in the parallel proceeding;
5. whether the petitioner and the defendant in the parallel proceeding are the same party; and
6. other circumstances that impact the Board’s exercise of discretion, including the merits.

Fintiv I, IPR2020-00019, Paper 11 at 5–6.

These factors are analyzed together to determine whether the “‘effect ... on the economy, the integrity of the patent system, the efficient administration of the Office, and the ability of the Office to timely complete proceedings,’ ... favors denying a petition even [if] some claims meet the threshold standards for institution under 35 U.S.C. §§ 314(a), and 324(a).” *See* Consolidated Trial Practice Guide (“CTPG”), 58 (quoting § 316(b)), available at <https://www.uspto.gov/Trial>

PracticeGuideConsolidated (Nov. 14, 2019). Here, none of the *Fintiv* factors weigh in favor granting institution.

A. Petitioner Has Not Requested a Stay Based on Its IPR Petition

The district court has not granted a stay, and Petitioner has not requested one based on its IPR Petition. *See* Pet. 70 (“neither party has moved to stay the ... Litigation.”). “In the absence of specific evidence, [the Board] will not attempt to predict how the district court in the related district court litigation will proceed because the court may determine whether or not to stay any individual case, ... based on a variety of circumstances and facts beyond our control and to which the Board is not privy.” *Sand Revolution v. Continental Intermodal Group-Trucking LLC*, IPR2019-01393, Paper 24 at 7 (PTAB Jun. 16, 2020) (informative). Petitioner does not present any “specific evidence” and, thus, cannot show that this factor weighs against discretionary denial under *Sand Revolution*. Pet., 69.

However, looking to similar cases, motions to stay when the PTAB has not yet instituted post-grant proceedings are consistently denied in the Eastern District of Texas, where the Litigation is pending. *Trover Group, Inc. v. Dedicated Micros USA*, No. 2:13-cv-1047-WCB, 2015 WL 1069179, at *6 (E.D. Tex. Mar. 11, 2015) (Bryson, J.) (“This Court’s survey of cases from the Eastern District of Texas shows that when the PTAB has not yet acted on a petition for *inter partes* review, the courts have uniformly denied motions for a stay.”). Further, Judge Gilstrap—before whom

the Litigation is pending—has denied motions to stay under similar circumstances. *See Samsung Elecs. Co. Ltd. v. Clear Imaging Research, LLC*, IPR2020-01402, Paper 12 at 12-13 (PTAB Mar. 3, 2021) (denying institution because, in part, Judge Gilstrap deemed unlikely to stay a case in similar circumstances). In fact, Judge Gilstrap granted less than 14% of motions to stay pending an IPR in 2023. *See* EX2003, 1-2 (denying 19/22 motions to stay pending IPR in 2023).

Thus, specific evidence suggest that the district court here would deny a motion to stay the Litigation. This first factor therefore favors denial of institution or, at best, is neutral.

B. The District Court Trial Will Occur Months Before the Board Could Reach a Final Written Decision in this Proceeding

Petitioner concedes “this factor weighs in favor of discretionary denial.” Pet., 70-71.

The Litigation is scheduled for trial in late 2024 (*see* §IV)—at least four months before the Board would reach a final written decision in mid April 2025 if this IPR were instituted. *Id.* Further, the median time-to-trial—which the Board may view as a more reliable indicator of when a trial in the Texas Litigation would most likely begin—would put the start of the trial roughly seven months before the Board would reach a final written decision. *Id.* This factor therefore strongly favors denial of institution. *Fintiv I*, at 9; *Fintiv II* at 13.

C. The District Court and Parties Have Invested and Will Invest Significantly in the Litigation

As of the date of this POPR, the parties have already invested resources and will invest significantly more resources by the time an institution decision is reached. *See generally* EX2004, EX1007.

In the Litigation, both parties have asserted infringement and invalidity of their respective patents. As such, the parties have already exchanged both preliminary and amended infringement and invalidity contentions. EX2004, 2-3. Petitioner concedes that they have also conducted a diligent prior art search prior to serving their invalidity contentions. Pet., 71. Numerous other document requests, interrogatories, and other discovery have already taken place, including: infringement contentions and supporting exhibits served by Patent Owner and amended infringement contentions with detailed source code allegations; infringement contentions served by Petitioner; 21 Interrogatories and 107 document requests served by Patent Owner, to which Petitioner has responded; 47 interrogatories served by Petitioner; invalidity contentions and amended invalidity contentions served by Petitioner, and invalidity contentions and amended invalidity contentions served by Patent Owner. *See* EX2004, 2-3.

In addition to thousands of pages of contentions, interrogatories, and document requests, the parties will have exchanged terms for claim construction, exchanged preliminary claim constructions and extrinsic evidence, and completed

the joint claim construction statement before the Board decides on institution. EX1007, 4. This will require completing and serving expert declarations pursuant to P.R. 4-3(b). Further, claim construction discovery will be underway before the institution decision is made and will be completed shortly after the decision. EX1007, 4. While no substantive orders have issued, the parties have still invested—and will continue to invest—significant resources into the litigation proceeding. Accordingly, this factor is neutral or should weigh at least slightly in favor of discretionary denial.

Further, as explained above in Section V.B, all aspects of the trial (e.g., fact discovery, expert discovery, dispositive motions, and the trial itself) will be completed before the Board’s final ruling on this matter. Thus, the parties will invest significant resources by the time the Board can reach a final decision.

D. The Petition and the Parallel Litigation Will Have Significant Overlap That is Not Mitigated by Petitioner’s Stipulation

1. Prior Art & Invalidity Theories

Petitioner relies on its stipulation to diminish the amount of overlap between the two parallel proceedings. Pet., 73. However, Petitioner deliberately pursues a stipulation that is significantly narrower than the *Sotera* stipulations typically given weight by the Board. See Interim Procedure for Discretionary Denials in AIA Post Grant Proceedings with Parallel District Court Litigation (“Guidance Memo”), 7. For example, in the precedential *Sotera Wireless* decision, the Petitioner stipulated

to “not pursue in [the District Court] ... any other ground ... that was raised or could have been reasonably raised in an IPR (*i.e.*, ***any ground that could be raised under §§ 102 or 103 on the basis of prior art patent or printed publications***).” *Sotera Wireless, Inc. v. Masimo Corporation*, IPR2020-01019, Paper 12 at 18-19 (PTAB Dec. 1, 2020) (precedential) (emphasis in original).

Here, Petitioner’s stipulation is only for “the same grounds” and leaves open the possibility to pursue any other combination or sub-combination based on the references cited in the Petition. Pet., 73. The stipulation also leaves open the possibility of Petitioner raising substantively identical invalidity theories based on prior art references that are substantially similar to the art in this proceeding. Such a stipulation fails to address the concerns of duplicative efforts and does not materially weigh against discretionary denial. Thus, the Board should exercise its discretion to deny institution here, where there is significant overlap of art in both proceedings and Petitioners have not proposed a stipulation as to any ground that reasonably could have been raised. *See, e.g., Samsung Elecs. Co. v. Ancora Techs., Inc.*, IPR2020-01184, Pap. 11 at 18-22 (P.T.A.B. Jan. 5, 2021) (finding narrower stipulation does not mitigate the concerns of inefficiency and determining the factors weights slightly in favor of exercising discretion to deny institution).

2. *Additional Claims*

Petitioner asserts that 1, 4, 7, 8, 9, 13, and 14 are at issue in the Litigation and that claims 1-20 are challenged in this petition. Pet., 72-73. Petitioner then concludes—without any explanation that this factor weighs strongly against denial. This is incorrect. It appears that Petitioner is attempting to argue that at least some of the challenged claims are not at issue in the Litigation. However, this does not weigh against discretionary denial.

Petitioner concedes independent claim 1 and dependent claims 4, 7-9, 13, and 14 are at issue in both proceedings. Pet., 72-73. Petitioner does not argue that the non-overlapping claims (e.g., 2-3, 5-6, 10-12, and 14-20) differ significantly in some meaningful way or that Petitioner would be harmed if institution of the non-overlapping claims is denied. In fact, Petitioner's challenges to independent claims 16 and 20 merely incorporate its analysis of claim 1, which confirms that Petitioner's challenges of the non-overlapping independent claims still overlap in subject matter. *See* Pet., 38-44, 64-66, 67-69. The same is true for at least some of the non-overlapping dependent claims. *See* Pet. 41, 66.

Accordingly, this factor thus weighs in favor of a discretionary denial. *Fintiv*, Paper 11 at 13 & 13 n.25 (citing *Next Caller, Inc. v. TRUSTID, Inc.*, IPR2019-00961, Paper 10 at 14 (PTAB Oct. 16, 2019) (denying institution, even though the petitions jointly involve all 52 claims of the patent and the district court parallel proceeding

involves only 7 claims, because the claims all are directed to the same subject matter and petitioner does not argue that the non-overlapping claims differ significantly in some way or argue that it would be harmed if institution of the non-overlapping claims is denied)).

Overall, because there will likely be overlap between the issues raised in the Petition and the parallel litigation, this factor weighs at least somewhat in favor of denying institution.

E. The Parties Are the Same in the Litigation

Petitioner concedes this “weighs in favor of discretionary denial.” Pet., 73.

F. The Petition Lacks Compelling Merits

None of the factors above weigh against discretionary denial. Petitioner concedes that factors 2 and 5 weigh in favor of discretionary denial and, for the reasons above, each of factors 1, 3, and 4 slightly weigh in favor of discretionary denial or are neutral. As a balancing of *Fintiv* factors 1–5 weighs in favor of denying institution, the Board should look to whether the Petition presents compelling evidence of unpatentability. *See Commscope Techs. LLC v. Dali Wireless, Inc.*, IPR2022-01242, Paper 23 at 4 (PTAB Feb. 27, 2023) (precedential).

A challenge does not present “compelling merits” unless the evidence of record at the institution stage is ***highly likely*** to lead to a conclusion that one or more claims are unpatentable by a preponderance of the evidence. *See OpenSky Indus.*,

LLC v. VLSI Tech. LLC, IPR2021-01064, Paper 102 at 49 (PTAB Oct. 4, 2022) (Director decision, precedential) (describing compelling merits as those that “plainly lead to a conclusion that one or more claims are unpatentable,” and noting that such standard can be met only “if it is highly likely that the petitioner would prevail with respect to at least one challenged claim”). “A compelling merits standard is a **higher standard** than the reasonable likelihood required for the institution of an IPR under 35 U.S.C. 314(a), and ... is also **higher** than the preponderance of the evidence standard (more likely than not) that applies to final determinations of patentability at the close of trial.” Advance Notice of Proposed Rulemaking Published April 21, 2023 (“Proposed Rulemaking”), at 16-17. “[O]nly a finding under this higher standard ... would compel the Board to review claims for the public benefit when other considerations favor discretionary denial.” *CommScope Techs. LLC v. Dali Wireless, Inc.*, IPR2022-01242, Paper 23, 3-4 (Feb. 27, 2023) (precedential).

Denial of institution under §314(a) is appropriate where Patent Owner raises an “argument [that] appears credible” that the alleged prior art did not disclose a limitation of the challenged claims. *See Kranos Corp. v. Apalone, Inc.*, IPR2020-00501, Pap. 13 at 14 (P.T.A.B. July 16, 2020); *Ericsson Inc. v. Collision Comm. Inc.*, IPR2022-01233, Paper 12 at 17-18 (P.T.A.B. January 19, 2023) (No compelling merits where PO presents viable arguments the Petitioner’s position lacks evidence). As explained below in Sections VI and VII, Petitioners’ arguments

depend on conclusory assertions and unconnected block quotations of the cited references, as well as misinterpretations of the teachings of those references. Thus, the grounds themselves do not meet the compelling merits standard that might otherwise support institution.

Further, Dr. Williams’ testimony—which repeats the Petition, *verbatim*—does not provide any additional probative evidence that a POSITA would have understood the cited portions of DaCosta and Kraft as teaching the claimed features. Instead, Dr. Williams mirrors the overly broad, conclusory assertions made in the Petition, which improperly shifts the burden to the Board to attempt to connect the claim language with the identified features of the cited references. *See Dynamic Drinkware, LLC v. Nat’l Graphics, Inc.*, 800 F.3d 1375 (Fed. Cir. 2015) (“In an IPR, petitioner has the ultimate burden of persuasion to prove unpatentability by a preponderance of the evidence, and this ***burden never shifts to patent owner.***”). Thus, Petitioner’s analysis does not provide evidence sufficient to “plainly” result in a conclusion of unpatentability and *Fintiv* factor 6 weighs in favor of discretionary denial.

At a minimum, factors 2, 4, 5 and 6 favor discretionary denial with factors 1 and 3 being neutral or slightly weighing in favor of discretionary denial. Balancing these factors based on the record shows a strong interest for the Board to exercise its discretion to deny institution. Accordingly, given the lack of strong countervailing

considerations, the improbability of a stay being granted, the investment of the parties, the overlap of issues, Petitioner’s decision to *not* include a *Sotera* stipulation, and the imminence of trial favor exercising discretion to deny institution in this proceeding.

VI. GROUND 1: THE PETITION FAILS TO SHOW ANTICIPATION OF CLAIMS 1, 16, AND 20 BY DACOSTA

“A claim is anticipated [under 35 U.S.C. § 102] only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference.” *Verdegaal Bros. Inc. v. Union Oil Co.*, 814 F.2d 628, 631 (Fed. Cir. 1987). Moreover, “[b]ecause the hallmark of anticipation is prior invention, the prior art reference—in order to anticipate under 35 U.S.C. § 102—must not only disclose all elements of the claim within the four corners of the document but must also disclose those elements ‘*arranged as in the claim.*’” *Net MoneyIN, Inc. v. VeriSign, Inc.*, 545 F.3d 1359, 1369 (Fed. Cir. 2008). Thus, invalidity for anticipation requires that “[t]he identical invention must be shown *in as complete detail* as contained in the patent claim.” *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 1236 (Fed. Cir. 1989); *see In re Skvorecz*, 580 F.3d 1262, 1266 (Fed. Cir. 2009) (“A rejection for ‘anticipation’ means that the invention is not new. Anticipation requires that all of the claim elements and their limitations are shown in a single prior art reference.”).

A. Claims 1, 16, and 20 are Novel Over DaCosta

Petitioner contends that claims 1-20 are anticipated by DaCosta. Pet., 14. Petitioner and Dr. Williams rely on DaCosta to expressly teach all limitations of the claims and do not assert any argument that such limitations are impliedly or inherently present in DaCosta. See Pet., 14-44; EX1003, ¶¶64-157. Thus, any inherency argument is waived, and all claim elements must be expressly disclosed by DaCosta for Petitioner to succeed. *Askeladden LLC v. Encoditech LLC*, IPR2017-00452, Paper 8, 21 (“Inherency must be alleged in the Petition for [the Board] to consider the issue.”); see *Pfizer Inc. v. Chugai Pharmaceutical Co. Ltd.*, IPR2017-01357, Paper 56, 18-19 (Nov. 28, 2018).

Independent claims 1, 16, and 20 include similar limitations. See §II.B. The Petition only analyzes the limitations of claim 1 (Limitations [1a]-[1j]) and then incorporates this analysis for claims 16 and 20. Pet., 38-41 (claim 16), 42-44 (claim 20). The only additional analysis provided for claims 16 and 20 is with respect to Limitation [16b], in which Petitioner contends that DaCosta’s navigation script and extraction script correspond to the claimed first and second sets of instructions, respectively. See Pet., 38-44. Thus, Petitioner’s analysis of claims 1, 16, and 20 can be treated as identical for purposes of the limitations addressed below, and the Petition’s deficiencies with respect to claim 1 must also apply to claims 16 and 20. Petitioner fails to establish DaCosta anticipates claims 1, 16, and 20.

1. ***[1g], [16g], [20g]: selecting the target database and target locations in the target database where the at least one of the data elements is to be stored in the target database***

The claims require “selecting the target database and target locations where the at least one of the data elements [selected from the list of data elements] are to be stored.” *Supra*, §II.B (Limitation [1e]).

To start, Petitioner identifies a user selecting text from a web page via the extraction script (e.g., 22) as corresponding to Limitation [1e]’s “selecting at least one of the data elements ...”. Pet., 18 (identifying “the user would then make a selection from the web page” at step 170 of FIG. 4 and “the user then highlights (i.e., using a mouse, drags over) the desired information”).

For the target database recited in Limitation [1g], the Petition identifies the “smart web server” depositing data into a knowledge repository and contends the smart server has the ability to use a distributed repository or a centralized repository. Pet., 20. Petitioner then contends that DaCosta discloses selecting a particular target database and location “through a navigation script.” Pet., 21 (citing EX1004, 5:12-20 and FIG. 2). Petitioner’s theory is unclear, inconsistent with its prior mapping, and based on a misguided reading of DaCosta. Almeroth, ¶68. (“Dr. Williams’ theory is inconsistent with both its prior mapping and the teachings of DaCosta”).

a) Petitioner's treatment of DaCosta's repositories is misguided

As explained in Section III.A, DaCosta's smart server operates in the background mode to fill a knowledge repository with data. EX1004, 18:8-12; *see* Almeroth, ¶¶58, 69. But this data is different from the data selected from the web page via the extraction module 20. *See* Almeroth, ¶¶59-60, 69; *supra*, §III.A.

When filling the knowledge repository with its smart server, DaCosta states:

Smart servers provide distributed, multiple dynamic virtual data feeds. The data feeds Web Folders that comprise a knowledge repository accessible by smart clients. *In an alternative embodiment* of the present invention, *extracted data can be stored* in a conventional repository, including Microsoft Exchange 2000 or any other database that constitutes a knowledge repository.

EX1004, 18:14-23. That is, DaCosta discloses two “alternative embodiment[s]” for this knowledge repository: (1) the *alleged* distributed repository, where the smart server “provides distributed, multiple dynamic virtual data feeds” to “Web Folders” and (2) the *alleged* conventional repository, where “data can be stored in a conventional repository.” EX1004, 18:14-23. While this knowledge repository stores *data* from the smart server, DaCosta discloses another embodiment with (3) a separate, centralized repository that stores *scripts*. EX1004, 18:34-39.

A POSITA would understand that DaCosta discloses two alternative server architectures for storing extracted data in its knowledge repository and that its scripts

may be stored in a separate, centralized repository. Almeroth, ¶70. However, Petitioner and Dr. Williams ignore these teachings, and instead concoct a theory that relies on mixing and matching the functionalities of the *different* servers and repositories of *different* embodiments to arrive at the claimed features. *See e.g.*, Pet., 20-21 (asserting DaCosta selects between the (1) “distributed repository” of the smart server or a (3) “centralized repository”), 23 (asserting the target repository is the alternative, (2) “conventional repository” of the smart server or the (3) “centralized repository.”); Almeroth, ¶71.

As explained below, none of these identified repositories are selected to store the data elements. Petitioner’s haphazard combination of these distinct repositories cannot establish anticipation. *Microsoft Corp. v. Biscotti, Inc.*, 878 F.3d 1052, 1069 (Fed. Cir. 2017) (“anticipation is not proven by multiple, distinct teachings that the artisan might somehow combine to achieve the claimed invention.”); *Net MoneyIn*, 545 F.3d at 1369 (“it is not enough [in an anticipation rejection] that the prior art reference discloses part of the claimed invention, which an ordinary artisan might supplement to make the whole, or that it includes multiple, distinct teachings that the artisan might somehow combine to achieve the claimed invention.”).

b) The user-selected text is not stored in DaCosta’s knowledge repository

As explained in Section III.A, the role of the “smart server is to fill a knowledge repository in *background mode*.” EX1004, 18:8-12., DaCosta does not

disclose the smart server storing data in the knowledge repository other than in this “background mode.” Almeroth, ¶72. Accordingly, DaCosta does not expressly disclose storing the user’s “selection from the web page” at step 170 of DaCosta (which Petitioner relies on for “selecting at least one of the data elements” in Limitation [1e]) in the knowledge repository. Further, as explained above, the selection of the text at step 170 during recording of the navigation script is unrelated to the smart server. *Supra*, §§III.A.2-3; *see* Almeroth, ¶58 (“the recording and playback of user generated navigation and extraction scripts is not directly related to the operation of the smart server”). Thus, Petitioner’s identification of the smart server or knowledge repository cannot anticipate the claims. And the fact that this user-selected text *might* be stored in a database does not disclose that the text is stored in *the* knowledge repository. *Net MoneyIN*, 545 F.3d at 1371 (“[D]ifferences between the prior art reference and a claimed invention, however slight, invoke the question of obviousness, not anticipation.”); *Wasica Finance GmbH v. Continental Automotive Systems Inc.*, 853 F.3d 1272, 1284 (Fed. Cir. 2017) (“[a]mbiguous references do not anticipate a claim.”).

Moreover, a POSITA would understand this “background” operation to be operating without user interaction. Almeroth, ¶73 (citing EX2005, 5/6; EX2006, 1/7). On the other hand, the scripts shown in FIGs. 2 and 4 (navigation and extractions scripts, respectively) requires a user to “interact with the loaded page by

clicking on or activating links or buttons, entering data and so on.” EX1004, 7:16-28; *see also id.* (“user-generated event 60”), Pet., 18 (“the user would then make a selection;” “the user then highlights ... the desired information;” and “a user can utilize a ‘find text’ command”). A POSITA would understand this selection of text as a “foreground” operation that requires the active input of a user. Almeroth, ¶74.

DaCosta does not disclose storing the user-selected text (the alleged data elements) in the knowledge repository (the alleged target database). Almeroth, ¶75; *see supra*, §§III.A.2-3. Instead, the user-selected text is output to the browser of the user. Almeroth, ¶75; EX1004, 5:12-20 (The user “view[s] the final output using a browser such as Microsoft Internet Explorer”); *see id.*, FIG. 10 (describing to “view output on your browser”).² Thus, Petitioner’s assertion that DaCosta’s smart server teaches the claimed feature is unsupported and inconsistent with DaCosta’s teachings. Almeroth, ¶75.

c) The user-selected text is not stored in DaCosta’s centralized repository or via the navigation script

Petitioner’s identification that DaCosta stores scripts in a centralized repository is irrelevant to the user-selected text that is extracted from the webpage.

² Although not alleged by Petitioner, a POSTIA would *not* understand sending data to a browser as disclosing the claimed feature. Almeroth, ¶75.

Almeroth, ¶77; *compare* EX1004, 9:56-65 (describing a user selecting data from the web page) *with* EX1004, 18:33-36 (“scripts are stored at a centralized repository”). The scripts are distinct from the data extracted from the webpage, and DaCosta fails to disclose storing the user-selected text (the alleged data elements (Pet., 18)) in the centralized repository (the alleged target database (Pet., 20)). Almeroth, ¶77. Thus, Petitioner’s anticipation theory is unsupported.

Petitioner’s assertion that “the navigation script also controls the storage of the extracted data in a repository, e.g., target database, as shown in element 90 of Figure 2” is also unsupported. Pet., 21-22; Almeroth, ¶78. As explained in above Section III.A.1, FIG. 2 discloses recording a navigation *script* and saving the *script* as an XML file. EX1004, 7:17-34. The saving of the navigation script does not teach storing the extracted data in a repository. Almeroth, ¶78. Petitioner cites to nothing in DaCosta to support this contention, instead relying on the *unsupported*, parroted testimony of its expert, which is entitled to no weight. *See* 37 C.F.R. § 42.65(a) (“Expert testimony that does not disclose *the underlying facts or data* on which the opinion is based is entitled to little or no weight.”). In fact, the *navigation* script does not even have access to the extracted data and, therefore, cannot control the storage of this data. Almeroth, ¶78; *see supra*, III.A.2 (extraction playback module 24 extracts the data as a text range 210).

For the reasons above, a POSITA would not understand Petitioner's citations to the centralized repository or navigation scripts as disclosing the claimed feature.

d) DaCosta does not select between repositories

Petitioner's contention that "DaCosta discloses selecting a particular target database and location through the use of scripts" (Pet. 21) is unsupported and inconsistent with the teachings of DaCosta. Petitioner submits no evidence to directly support this assertion. *See* Pet., 21 (failing to cite to its expert or DaCosta). This amounts to nothing more than attorney argument, which is entitled to no weight. *See Icon Health & Fitness, Inc. v. Strava, Inc.*, 849 F.3d 1034, 1043 (Fed. Cir. 2017).³ The only citation Petitioner provides for this argument is to the generalized description (EX1004, 5:12-20), which fails to disclose anything about selecting a repository or storing data within a repository. Almeroth, ¶79. In fact, this generalized description teaches "view[ing] the final output using a browser such as Microsoft Internet Explorer." EX1004, 5:12-20. Yet, sending information to a

³ Dr. Williams' parroting of this conclusory assertion without citing to any supporting evidence (EX1003, ¶81) is similarly entitled to no weight. *See* 37 C.F.R. § 42.65(a) ("Expert testimony that does not disclose *the underlying facts or data* on which the opinion is based is entitled to little or no weight.").

browser does not teach selecting the target database and target locations in the target database. Almeroth, ¶79.

Further, to the extent Petitioner is mapping DaCosta's disclosure of two alternative embodiments for the knowledge repository as *selecting* the repository, this is incorrect. The disclosure of two alternative embodiments does not disclose a selection between two repositories (e.g., databases). *Net MoneyIn*, 545 F.3d at 1369; Almeroth, ¶76. A POSITA would have understood that the "alternative embodiment" of a conventional repository is discussing how the server architecture would be initially configured. Almeroth, ¶76; *see* EX1004, FIG. 22 (illustrating "an architecture" for a smart server). In other words, the smart server would store data in only one of either the distributed repository or the conventional repository, depending on how the server was implemented during construction. Almeroth, ¶76 (citing EX1004, 5:1-3, 18:16-23, FIG.22). DaCosta does not disclose that the smart server would select between coexisting distributed and conventional repositories, and a POSITA would not understand the two "alternative embodiments" of the smart server as teaching the claimed feature. Almeroth, ¶76.

e) Petitioner's Theory is not sufficiently explained

Petitioner's analysis of Limitation [1g] cites to the smart server's knowledge repository (Pet., 20-21 (citing EX1004, 18:15-23)), the centralized repository for storing scripts, (Pet., 21 (citing EX1004, 18:33-38)), a generalized description of the

various embodiments of the invention (Pet., 21 (citing EX1004, 5:12-20)), and step 90 of FIG. 2 (Pet., 21-22 (not citing EX1004)). Yet, the centralized repository and knowledge repository are distinct repositories for storing different types of data. *See* Almeroth, ¶¶70-71; *supra*, §VI.A.1. Additionally, the smart server and the navigation script are distinct components that perform different processes. Almeroth, ¶60; *supra*, §VI.A.1.b). Petitioner does not explain how any of its quoted portions of DaCosta are related.

Moreover, none of these components are connected to the data selected by the extraction script, which Petitioner identifies as the “at least one of the data elements is to be stored in the target database.” Pet., 18-19 (identifying step 170 of FIG. 4 as disclosing Limitation [1e]). Petitioner’s piecemeal theory attempts to combine the distinct teachings of the smart server, navigation scripts, extraction scripts, knowledge repository, and centralized repository into some re-imagined process that is not disclosed or taught by DaCosta. But Petitioner’s unexplained ‘picking and choosing’ of disparate features of DaCosta cannot establish anticipation. *See Net MoneyIN*, 545 F.3d at 1371 (quoting *In re Arkley*, 59 CCPA 804, 455 F.2d 586, 587 (1972)); *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383 (Fed. Cir. 2001). And it is neither Patent Owner nor the Board’s burden to sift through Petitioner’s disparate citations to identify some undeveloped theory. *see* 35 U.S.C. § 322(a); *Harmonic v. Avid Tech*, 815 F.3d 1356, 1363 (“[T]he petitioner has the

burden from the onset to show with particularity why the patent ... is unpatentable.”).

2. ***[1i], [16i], [20i] – “setting a schedule frequency for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database”***

Petitioner cites to various portions of DaCosta as allegedly disclosing this limitation. Petitioner’s mappings both lack particularity and are substantively deficient for the reasons explained below.

a) Petitioner’s Theory is not sufficiently explained

Petitioner cites to three distinct aspects of DaCosta for this limitation without explanation as to whether these three aspects are somehow connected or if they are alternative theories. *See* Pet., 24-25 (citing to generalized description, form parameters, and looping of smart server). Petitioner also fails to explain how these aspects relate to the previous mappings of the claimed features, which cite to additional, distinct aspects of DaCosta. This layering of arguments without explanation of how possible alternatives interrelate cannot meet the particularity requirement. *See U.S. Well Services*, IPR2022-01163, Paper 9, 15-17. Additionally, this approach “improperly shifts the burden of deciphering Petitioner’s arguments onto Patent Owner and the Board, which is contrary to the statutory and regulatory provisions governing *inter partes* review proceedings.” *See id.* For at least this reason, Ground 1 must fail.

b) DaCosta does not set a schedule frequency for transferring data elements to the target database

Turning now to Petitioner's theories, the Petition first cites to a generalized description (EX1004, 5:12-20) that discloses DaCosta's invention can enable a user to "automatically navigate to a plurality of web site[s]," "extract specified information," "process the extracted data according to customizable scripts, integrate information from other applications such as Microsoft ('MS') Word," and "view the final output using a browser." Pet., 24. DaCosta states that these steps may be repeated in a "scheduled manner." EX1004, 5:12-20.

However, none of the steps identified by petitioner disclose transferring data to a database. Almeroth, ¶81; EX1004, 5:12-20 (disclosing to "view the final output using a browser" instead of transferring to a database); *see supra*, §VI.A.1.b). Thus, it is irrelevant that these steps are repeated in a "scheduled manner" as they do not disclose transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database. Almeroth, ¶81; *see also* Pet., 24 (alleging repeating of the steps of "extracting, processing, [and] integrating," but **omitting** transferring to a database).

Second, Petitioner's mapping to DaCosta's form parameters is misplaced. Pet., 24-25 (citing EX1004, 7:62-65); Almeroth, ¶82. Petitioner's assertion that "DaCosta discloses the use of parameters to control setting a schedule frequency" is unfounded. Almeroth, ¶82. The form parameters relate to the form fields in the web

page, and form variations can be used to change the data entered into the form fields. Almeroth, ¶82 (citing EX1004, 7:29-31, 55-64); *see also* Almeroth, ¶54; EX1004, 7:47-51 (navigation script records the “form fields” that were acted upon). By using variations of the information originally input into the form fields, the system can access similar webpages by changing the originally recorded form field inputs into new inputs. EX1004, 7:55-64 (playback module performs form variations “specifying new form parameters to replace the ones originally recorded.”).

DaCosta provides an example of this form variation process where “a form can be filled in to get pricing information from an online bookstore for different titles by changing a single parameter in a query form” (e.g., the ‘title’ form field). Almeroth, ¶83 (quoting EX1004, 7:65-8:3). That is, once an original navigation script is recorded for a single tile from the online bookstore, the playback module can input multiple form variations (e.g., different titles of books) into the form field to access information on multiple books. Almeroth, ¶83. In this way, the script can be modified to run multiple queries “based upon a single exemplomatic query” instead of the user having to create new scripts to run separate queries for every variation. EX1004, 8:3-5; Almeroth, ¶83. But these variations have nothing to do with setting a schedule frequency. Almeroth, ¶83.

Finally, Petitioner’s citation to the smart server looping “after a *preset* delay” does not disclose setting a schedule frequency. EX1004, 18:23-25; Almeroth, ¶84.

As explained above in Section VI.A.1.b), the user-selected text that Petitioner identifies as the “at least one of the data elements” in Limitation [1e] is not stored in the knowledge repository or utilized by the smart server. Here, Limitation [1i] requires setting a schedule frequency *for transferring* to the target database the corresponding “data elements” that were selected in Limitation [1e]. *Supra*, §§III.A.3, VI.A.1.a). Thus, for the same reasons explained above, the operation of the smart server cannot disclose the claimed feature. Almeroth, ¶84 (the smart server is a distinct embodiment that stores data in a background mode and is unrelated to the data elements identified by Petitioner).

Independently, a POSITA would not understand a “*preset* delay” as disclosing setting a schedule frequency. Almeroth, ¶84 (the fact that a delay was previously set does not disclose setting the delay); EX2007, 1 (preset means “set in advance”).

3. *[1h], [16h], [20h] “saving the source locations and target locations of the at least one of the parsed data elements in a location file in the target database;”*

Petitioner identifies the URL or metabookmarks as corresponding to the source location. Pet., 19 (Limitation [1f]). Petitioner then identifies two distinct repositories—the knowledge repository for *data*, and centralized repository for *scripts*—as corresponding to the target location and target database. Pet., 20-21; *supra*, §VI.A.1.

Petitioner’s evaluation of Limitation [1h] introduces further inconsistencies that diverge from its previous mappings of the URL or metabookmarks (the alleged source location) or the target locations (which Petitioner failed to identify). *See* Almeroth, ¶85. In addition to these inconsistencies, Petitioner’s analysis is largely unexplained and based on a misinterpretation of DaCosta.

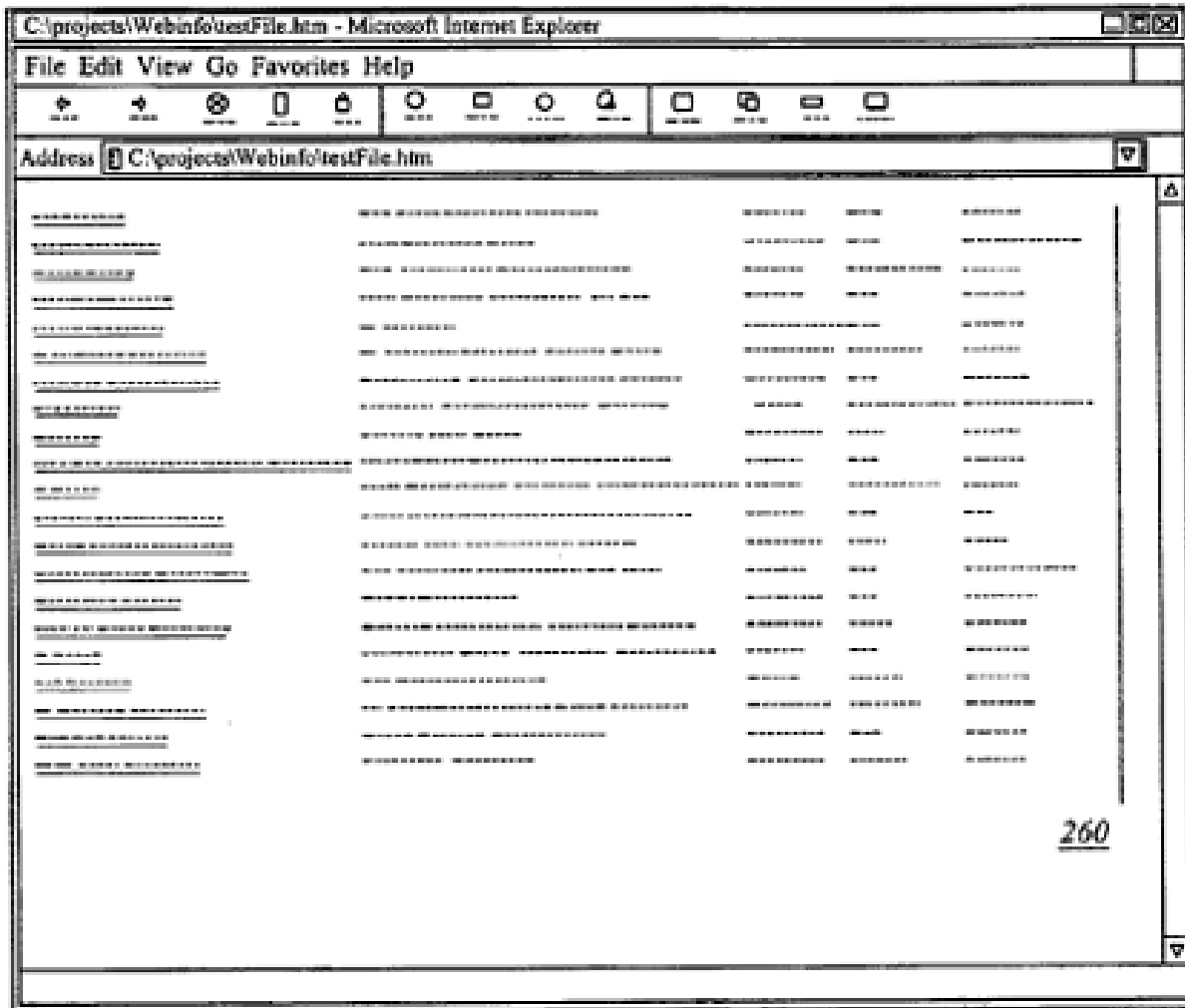
a) The Navex scripts do not disclose a target location in a target database

Petitioner contends that DaCosta’s navigation and extraction scripts (Navex scripts) “include source and destination information.” Pet., 22. Petitioner’s citation to Dr. Williams—which repeats the Petition’s assertions verbatim—and FIG. 10 of DaCosta fail to provide any evidentiary support to Petitioner’s theory.

Referring to DaCosta’s FIG. 10 (inset below), the Navex scripts can “extract numerous records 260” based on a pattern that was derived by the extraction recording module 22.

Your JobFinder Output

Save your "script" and run it on demand. View output on your Browser.
Save it in Excel, Access or Word. Do it Every Hour - automatically.



Automatic extraction of jobs data. Data may also be transferred to a data base for SQL processing

FIG. 10

EX1004, 11:19-24. A POSITA would not understand the identification of this pattern, or the extraction of data based on the pattern, to disclose "destination information" or a target location in the target database. Almeroth, ¶87. The pattern

tells you which information to extract from a webpage, not where to send or store that information. Almeroth, ¶87; EX1004, 9:7-14 (“a pattern according to which data can be automatically extracted from a web page”). To illustrate, the data 260 from FIG. 10 that matches the pattern is “output on your browser.” EX1004, FIG. 10. Thus, the Navex script that Petitioner cites to sends the data to a browser, not the target databases identified by Petitioner. *See* Pet., 20; Almeroth, ¶87.

Petitioner also quotes the statement that “data may also be transferred to a data base for SQL processing.” However, as explained in Sections VI.A.1.b)-c) (above) and Section VI.A.3.b) (below), DaCosta does not disclose that this data would be transferred to either the knowledge repository for background *data* or the centralized repository for *scripts* identified by Petitioner. And, to the extent Petitioner intends to assert that the browser or the database for SQL processing is the same as the knowledge repository, this is neither taught nor disclosed in DaCosta. Almeroth, ¶88. And, to the extent the Board might consider trying to read Petitioner’s disparate citations together, the result would be *at best* an ambiguous mashup that still could not support anticipation. *Wasica Finance*, 853 F.3d 1272, 1284 (“[a]mbiguous references do not anticipate a claim.”).

b) Petitioner’s mapping of the Databases is inconsistent with DaCosta

Petitioner’s analysis of the conventional, centralized, and decentralized, repositories is incompatible with DaCosta’s teachings. Almeroth, ¶89; *see supra*,

§§III.A, VI.A.1.a). As explained above, DaCosta discloses a knowledge repository that is filled with extracted data by “a smart server” operating in “background mode.” EX1004, 18:10-12. This knowledge repository may take the form of: (1) multiple dynamic virtual data feeds—the distributed repository—or, alternatively, (2) a conventional repository. Almeroth, ¶89; EX1004, 18:14-23; *see also* Almeroth, ¶70. In addition to the knowledge repository—which stores extracted data—a preferred embodiment of DaCosta can store scripts in (3) a centralized repository. EX1004, 18:33-36.

Petitioner’s assertion that the target database “may be a conventional repository or a centralized repository” is wrong and attempts to mix different distinct repositories to gloss over DaCosta’s differences. Almeroth, ¶90 (Dr. Williams improperly misconstrues the functions of the two separate databases to create something that is not disclosed in DaCosta). The conventional repository is for storing extracted data from the smart server, and the centralized repository is for storing scripts. Almeroth, ¶90 (citing EX1004, 18:8-23, 34-40); *see supra*, §VI.A.1.a). Petitioner’s contention that “scripts may also be stored in a decentralized target database” (Pet., 24) is similarly misplaced. Almeroth, ¶¶90-91. In fact, this contention is nonsensical when viewed next to the quoted portion of DaCosta (EX1004, 18:17-23), which states that “*extracted data* can be stored in a *conventional repository*”—**not** that *scripts* can be stored in a *decentralized* database.

4. [16b], [16j]: First and second sets of instruction codes

Petitioner contends that DaCosta's navigation script and extractions script correspond to the claimed first and second sets of instruction codes. Pet., 39 (“a navigation script e.g., a first set of instructions, and an extraction script, e.g., a second set of instructions”). Petitioner then incorporates its analysis of claim 1 for the remaining limitations of claim 16 ([16c]-[16j]). Pet., 39-41. But this incorporation creates internal inconsistencies that are fatal to Ground 1.

The claimed first set of instruction codes are “for initializing a data retrieval stage by: [performing the steps of Limitations [16c]-[16i].” Petitioner incorporates its respective analysis of Limitations [1c]-[1i] for Limitations [16c]-[16i]. However, Petitioner relies on the extraction script—the alleged second set of instruction codes—for teaching several different elements of Limitations [1c]-[1i]. In fact, the Petitioner cites to the extraction script (or extraction module 20) as disclosing portions of Limitations [1c], [1d], [1e], [1g], and [1h]—relying exclusively on the extraction module for Limitation [1e]. Pet., 15-19 (citing to the “extraction module,” “extraction scripts,” and citing to the steps of “Recording Extraction Script” as illustrated in “FIG. 4”) 21-23 (discussing “extraction scripts”). Petitioner's reliance of the navigation script while incorporating discussion of the extraction script renders the Petitioner internally inconsistent. This is fatal to Ground 1. *See e.g., Ex Parte Obrien*, Appeal No. 2002-1846, 2002 WL 33950052, at *2 (BPAI Sept. 20,

2002) (non-binding) (reversing rejections that were “not adequately articulated and appear[ed] *internally inconsistent*”).

Petitioner analysis of the second set of instruction codes also includes similar inconsistencies. *See* Pet., 25 (citing to the *navigation* “playback module 14” and “*navigation* recording module 12” for allegedly teaching aspects of Limitation [1j]), 61 (identifying the “*extraction* script” as the second set of instructions and incorporating the discussion from Limitation [1j])

VII. GROUND 2: THE PETITION FAILS TO SHOW ANTICIPATION OF CLAIMS 1, 16, AND 20 BY KRAFT

A. Claims 1, 16, and 20 are Novel Over Kraft

Petitioner contends that claims 1-20 are anticipated by Kraft. Pet., 44-69. Petitioner does not assert any inherency arguments and, therefore, such arguments are waived. *Askeladden*, IPR2017-00452, Paper 8, 21.

Petitioner’s analysis of claims 16 and 20 is the exact same as that for claim 1. *See* Pet., 64-69 (incorporating “the above discussion of claim 1” for every limitation of claims 16 and 20 with no independent analysis). Thus, Petitioner’s analysis of claims 1, 16, and 20 can be treated as identical for the below distinctions, and the Petition’s deficiencies with respect to claim 1 also apply to claims 16 and 20. Petitioner thus fails to establish that Kraft anticipates claims 1-20.

1. [16b], [16j]: First and second sets of instruction codes

Petitioner fails to address the claimed first and second sets of instructions that are recited in claim 16. Petitioner provides no analysis for these claimed features and merely “incorporates by reference the above discussion of claim 1” and cites to paragraphs 223 and 231 of Dr. Williams’ declaration. Pet., 64-66. However, the Petition fails to identify any “set of instruction code” in its analysis of claim 1 (Pet., 44-56), and the cited paragraphs of Dr. Williams declaration provide only unsupported, conclusory assertions. As the petition fails to identify anything in Kraft as allegedly disclosing the claimed first and second sets of instruction codes, the Petition cannot establish anticipation of claim 16. *Verdegaal Bros*, 814 F.2d at 631 (anticipation requires disclosure of “*each and every element* as set forth in the claim”).

2. [1i], [16i], [20i]: – “setting a schedule frequency for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database;”

Petitioner provides three theories for this claimed feature—all of which are deficient. Almeroth, ¶96.

First, Petitioner alleges that “Kraft discloses setting a schedule frequency where ‘users will be able to return to the source document, and optionally to use applications for automatically synchronizing a destination document to the target document.’” Pet., 54 (quoting EX1005, ¶[0009]). A user returning to a source

document (e.g., web page) fails to disclose setting a schedule frequency. Almeroth, ¶97; *Cf* EX1001, 5:17-19 (“the user sets the timer 350 by selecting the frequency of data retrieval; i.e., every 5 minutes, 15 minutes, at a specific time such as 12:00 p.m., etc.). Additionally, the use of applications for automatically synchronizing a destination document to the target document does not disclose setting a schedule frequency. Almeroth, ¶98. This synchronization application is based on “the user [requesting] to update the destination document,” not setting a schedule frequency. Almeroth, ¶98 (quoting EX1005, ¶[0016]). To the extent that Petitioner is relying on the term “automatically,” this term is simply referring to the user not having to do anything *after* requesting for the destination document to be updated. Almeroth, ¶98. Kraft plainly does not disclose that this synchronization application sets a schedule frequency and, therefore, the synchronization application cannot anticipate the claimed feature.

Next, Petitioner cites that Kraft’s “system 10, transparently to the user, continuously or periodically operates in the background.” Pet., 54 (quoting EX1005, ¶[0044]). This quoted portion fails to disclose setting a schedule frequency. Almeroth, ¶99. The disclosure that a computing system may generally “operate” in the background does not teach the claimed features. Almeroth, ¶99.

Finally, Petitioner cites to “the background section” of Kraft that states “[t]he crawler is run periodically to update previously stored data and to append

information about newly created web pages” and that “[t]he information compiled by the crawler is stored in a Metadata repository or database.” This too does not disclose setting a schedule frequency (Almeroth, ¶100), much less for transferring the at least one of the parsed data elements from the source locations to the target locations in the location file in the target database.

Further, Petitioner’s reliance on the operation of a crawler and Metadata repository from Kraft’s “background section” cannot establish anticipation. *TCT Mobile (US) Inc. v. Semiconductor Energy Laboratory Co., Ltd.*, IPR2021-01573, Paper 13 at 8 (April 26, 2022) (relying on portions of disclosed embodiment and embodiment “mentioned in [the] background discussion ... is not appropriate in an anticipation analysis”) (citing *Net MoneyIn*, 545 F.3d at 1369–71). Petitioner has not provided a link between Kraft’s embodiments and the crawler or Metadata repository in the background. Nor is there such a link described in Kraft. Rather, Petitioner has simply selected various disparate features from the references in arriving at the claimed invention which is improper under §102. *TCT Mobile*, at 8.

Additionally, none of Petitioner’s previous mappings relate to the crawler or Metadata repository. Pet., 45-54. Thus, to the extent that Petitioner’s reliance on the background disclosure could even be proper, it is inconsistent with the rest of Petitioner’s mapping. Petitioner maps to destination documents repository 170 as the claimed target database. Pet., 51 (“Kraft discloses the target database as the

‘destination documents repository 170’”). Any discussion of the crawler periodically updating data stored in a metadata repository is unrelated to documents repository 170. *Cf* EX1005, ¶[0053] (“crawler 150 ... downloads web documents to the search engine repository 210”). As such, the cited language in Kraft does not teach the claimed feature. *See* Almeroth, ¶100 (Dr. Williams’ theory is inconsistent with the teachings of Kraft, which does not disclose the claimed feature)

3. [1e], [16e], [20e]: – “selecting at least one of the data elements from the list to be stored in the target database”

Petitioner cites solely to paragraph [0059] of Kraft as allegedly disclosing this feature. Pet., 49. Petitioner offers no explanation of how this disclosure teaches the claimed feature, and merely quotes portions of paragraph [0059]. *Id.* A POSITA would not recognize paragraph [0059] as teaching the claimed limitation. Almeroth, ¶101.

Kraft discloses navigating to a target document 60 (e.g., a web page) and then saving the entire document in the destination documents repository 170. EX1005, ¶[0059]. There is no teaching of selecting data elements and, therefore, the cited portion of Kraft cannot disclose “selecting at least one of the data elements from the list to be stored in the target database.” Almeroth, ¶102.

To the extent that Petitioner intends to argue that the navigation to the target document discloses selecting at least one data elements, this is unsupported and inconsistent with the understanding of a POSITA. Almeroth, ¶103. The ’667 Patent

clearly distinguishes between selecting a web page (e.g., source file) and selecting data elements. EX1001, 2:20-22 (system “allows the user to [(1)] select one or more Web pages containing data of interest; [and (2)] to specify exactly which data within any page is to be captured”); Almeroth, ¶103 (citing EX1001, 4:44-55 (describing selecting a web page at step 410), 5:12-13 (describing selecting data in step 430)). Thus, the quoted portion of Kraft does not disclose the claimed feature. Almeroth, ¶103. For these reasons, Ground 2 fails.

4. [1d], [16d], [20d]: “parsing selected data elements in the source file into a list of data elements;”

Petitioner cites to two paragraphs of Kraft as allegedly disclosing the claimed feature. Neither paragraph teaches the limitation. Almeroth, ¶104. In fact, the disclosures of the identified Kraft paragraphs are also found in the ’667 Patent and are distinct from the disclosure of the claimed feature. *Compare* EX1005, ¶¶[0028], [0058]) *with* EX1001, 3:13-22, 4:6-16.

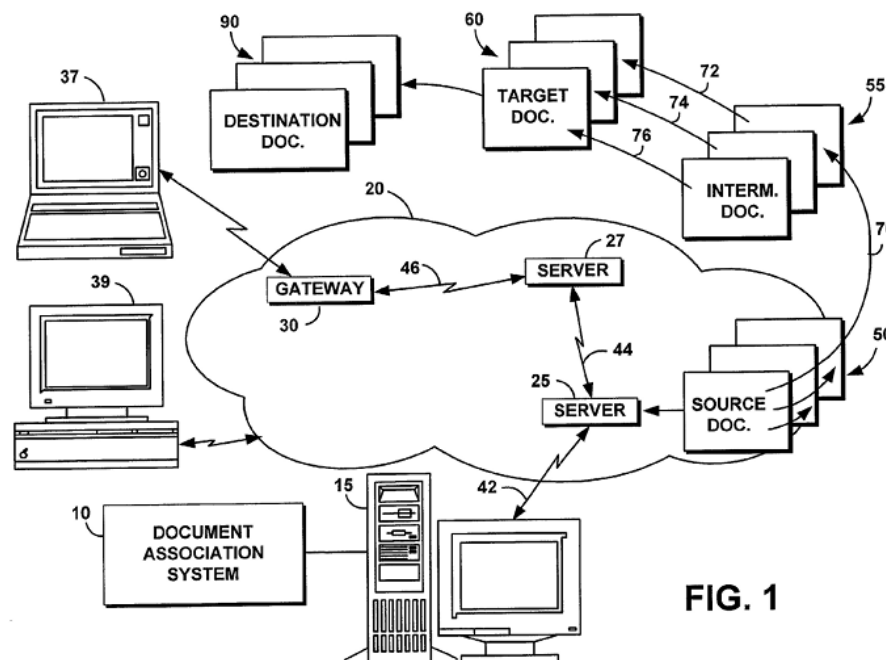
Petitioner’s first theory cites to the definition of “HTML (Hypertext Markup Language).” Pet., 48; EX1005, ¶[0028]. This definition “provide[s] background information pertaining to the technical field of the present invention.” EX1005, ¶[0025]. Specifically, this discloses that HTML is used for including attributes within web pages by the owner of the web page. *Id.*; Almeroth, ¶105. These attributes will be interpreted by a user’s web browser to display the document. *Id.*

A POSITA would not understand this background definition as disclosing the claimed feature. To start, this exact same definition is provided in the '667 Patent. EX1001, 3:13-22 (providing same definition of HTML).⁴ A POSITA—viewing the specification—would understand this background definition of HTML to be distinct from the claimed feature. Almeroth, ¶106. Petitioner provides no analysis of the claim and appears to rely on this definition because it is the only paragraph in Kraft that includes the term “parse.” *See generally*, EX1005. However, merely identifying a matching term from the claim language in a background definition is not sufficient for anticipation. *See Net MoneyIn*, 545 F.3d at 1369–71 (explaining that anticipation requires a clear disclosure of the claimed invention with elements arranged as in the claim and without any need for combining distinct disclosures). The fact that webpages are “interpreted by the browser and used to parse and display the document” (EX1005, ¶[0028]) does not disclose the claimed step of “parsing selected data elements in the source file into a list of data elements.” Almeroth, ¶106 (citing EX1001, 4:6-26, 4:46-59).

⁴ In fact, this “background definition” is provided in several patent applications filed by International Business Machines Corp (IBM)—the Applicant of the '667 Patent and Kraft—around this time. *See e.g.*, Sundaresan, 5:61-6:5; Ehrlich, 6:26-36; Miner, 6:6-16; Dill, 3:54-64.

Next, Petitioner contends that Kraft “discloses that the list of data elements may include such hyperlinks.” Pet., 48 (citing EX1005, ¶[0058]). At base, it is unclear what “list of data elements” Petitioner is referring to as Petitioner has failed to mention any “list,” “data,” or “elements” before this conclusion. Further, Petitioner does not explain how or why a POSITA would understand a website having hyperlinks as disclosing the claimed feature. This alone renders Petitioner’s analysis deficient for lacking sufficient particularity. Nevertheless, a POSITA would not understand the use of hyperlinks as disclosing this claimed feature. Almeroth, ¶107 (citing EX1001, 4:6-26, 4:46-59).

Steps 320-330 of FIG. 3 described navigating between web pages using hyperlinks, such as going from source document 50 to documents 55, as illustrated in Kraft’s FIG. 1 (inset below).



EX1005, ¶[0058]. This is the same type of navigation disclosed in the '667 Patent, where “servers 25, 27 provide access to stored information such as hypertext or Web documents indicated generally at 50, 55, and 60 ... [that include] hypertext links 70, 72, 74, 76 to other webs sites or documents 55, 60.” EX1001, 4:9-15. Yet, the '667 Patent distinguishes between navigating to an HTML page (e.g., page 220 of FIG. 2), and parsing selected data elements in the source file into a list of data elements. EX1001, 4:44-66 (selecting web page 220 in step 410 and parsing the source file in step 420). Almeroth, ¶108. In fact, a POSITA reading the '667 Patent would understand that the navigation to a desired HTML page (e.g., 220) would correspond to selecting the source file. Almeroth, ¶108. Petitioner itself cites to this same disclosure (EX1005, ¶[0058]) as disclosing Limitation [1c]. *See* Pet., 46-47.

Outside of the quoted portions of Kraft, Petitioner and its expert provide no analysis of the claimed feature. A POSITA would not understand either of the quoted portions of Kraft as disclosing the claimed feature and, therefore, the Petition cannot legally establish anticipation. Almeroth, ¶109.

VIII. CONCLUSION

For the reasons explained above, the Board should deny institution because Petitioner has not established any reasonable likelihood of success on any of its proposed grounds.

Dated: January 17, 2024

Respectfully submitted,

/Eagle H. Robinson/
Eagle H. Robinson
Reg. No. 61,361

Lead Counsel for Patent Owner

CERTIFICATE OF WORD COUNT

Pursuant to 37 C.F.R. § 42.24(a)(i), the undersigned certifies that this Paper—exclusive of the table of contents, table of authorities, certificate of service, and this certificate of word count—includes 11,627 words. The undersigned relies upon the word count generated by Microsoft Word.

/Eagle H. Robinson/
Eagle H. Robinson (Reg. No. 61,361)

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. § 42.6(e), the undersigned certifies that on January 17, 2024, complete copies of Patent Owner's Preliminary Response and EX2001-EX2011 were served on Petitioner via email (by consent), at the following email address(s):

tomd@cherianllp.com

rmolitors@cherianllp.com

elizabetho@cherianllp.com

/Eagle H. Robinson/

Eagle H. Robinson, Reg. No. 61,361