

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

BRIGHT DATA LTD.,
Petitioner,

v.

OXYLABS, UAB,
Patent Owner.¹

IPR2023-01442
Patent 7,020,667 B2

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

GIANNETTI, *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 II.A, n.2.

I. INTRODUCTION

Bright Data Ltd. (“Petitioner”) filed a Petition (Paper 2, “Pet.”) requesting *inter partes* review of claims 1–20 (the “challenged claims”) of U.S. Patent No. 7,020,667 B2 (Ex. 1001, “the ’667 patent”). Oxylabs, UAB (“Patent Owner”)² filed a Preliminary Response (Paper 6, “Prelim. Resp.”).

The Board has authority to determine whether to institute *inter partes* review. *See* 35 U.S.C. § 314; 37 C.F.R. § 42.4(a). Under 35 U.S.C. § 314(a), we may not authorize an *inter partes* review unless the information in the petition and the preliminary response “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 any claim challenged in the Petition. In addition, for the reasons stated on the merits of the Petition, we do not reach the question of whether to exercise our discretion to deny institution under 35 U.S.C. § 314(a). We, therefore, deny institution of *inter partes* review.

II. BACKGROUND

A. Related Matters

The parties identify the following district court proceeding involving the ’667 patent: *Bright Data Ltd. v. Oxylabs, UAB*, 2:23-cv-00171 (E.D. Tex.) Pet. 1–2; Prelim. Resp. 17.

² As set forth below, the ’667 patent is now owned by oxylabs, UAB. The caption of this proceeding has been revised to reflect this change.

B. Real Parties-in-Interest

Petitioner identifies Bright Data Ltd. as the real party-in-interest.

Pet. 1.

Patent Owner identifies the following real parties-in-interest: oxylabs, UAB, as successor in interest to metacluster lt, UAB and the owner of the '667 patent. Paper 4,1. Patent Owner states also that oxysales, UAB and metacluster lt, UAB have been merged into Teso LT, UAB, which now operates as oxylabs, UAB. *Id.*

C. The '667 Patent (Ex. 1001)

The '667 patent is titled “System and Method for Data Retrieval and Collection in a Structured Format.” Ex. 1001, (54). The '667 patent relates to “a software system and associated method for use with Web pages to transfer data from a Web page source file to a target database.” *Id.* at 1:8–10. More particularly, the patent discloses “a method for parsing a Web page source file, collecting the specified data on the Web page, automatically retrieving the specified data on a predetermined frequency or scheduled time, and then storing the collected data in a database structure.” *Id.* at 1:11–15.

The data collection and retrieval method and system disclosed by the '667 patent allow capture and replication of data presented at various Web pages into a database application through text parsing of the HTML source code of that document. *Id.* at (57). The system allows the user to select one or more Web pages containing data of interest; specify which data within any page is to be captured; specify how frequently data is to be collected; and specify the conditions for collection and retrieval. *Id.*

D. Illustrative Claim

The Petition challenges claims 1–20, of which claims 1, 16, and 20 are independent. Claim 1 is illustrative of the claimed subject matter and is reproduced below:³

1. [a] 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:

[b] initializing a data retrieval stage by:

[c] selecting the source file;

[d] parsing selected data elements in the source file into a list of data elements;

[e] selecting at least one of the data elements from the list to be stored in the target database;

[f] identifying source locations of the selected at least one of the data elements;

[g] 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;

[h] 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] 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

[j] 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

³ References added appear in square brackets. *See* Pet. 14–25.

the saved source locations in the source file to the saved target locations in the target database.

Ex. 1001, 6:63–7:24.

E. Prior Art References and Other Evidence

Petitioner relies on the following references (Pet. 4, 11, 13):

1. U.S. Patent 6,826,553 B1 (Nov. 30, 2004) (Ex. 1004, “DaCosta”); and
2. U.S. Patent Application Publication 2003/0018669 A1 (Jan. 23, 2003) (Ex. 1005, “Kraft”).

In addition to these references, Petitioner relies on a Declaration of Dr. Tim Arthur Williams. Ex. 1003 (“Williams Decl.”). Patent Owner has submitted a declaration of Kevin Almeroth, Ph.D. Ex. 2001 (“Almeroth Decl.”).

F. The Asserted Grounds of Unpatentability

Petitioner asserts the following two grounds of unpatentability.

Claim(s) Challenged	35 U.S.C. §⁴	Reference(s)
1–20	102	DaCosta
1–11 and 13–20	102	Kraft

Pet. 3–4.

⁴ Because the earliest application from which the ’667 patent issued was filed before March 16, 2013, the pre-AIA (“America Invents Act”) version of § 102 applies. Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284, 285–88 (2011).

III. ANALYSIS OF THE CHALLENGED CLAIMS

A. *Introduction*

As discussed, the Petition challenges patentability of the '667 patent claims on two separate anticipation grounds. Pet. 4. The first (identified by Petitioner as Ground 1) asserts that DaCosta anticipates claims 1–20. The second (identified as Ground 2) asserts that Kraft anticipates claims 1–11 and 13–20.⁵ *Id.*

Patent Owner's Preliminary Response requests that the Board exercise its discretion to deny institution in light of parallel litigation in the Eastern District of Texas involving the same parties in which the validity of the '667 patent is at issue. *See* Prelim. Resp. 17–29. The Preliminary Response also challenges Petitioner's anticipation analysis for each ground. *Id.* at 29–49 (first ground), 49–57 (second ground).

For each anticipation ground, Patent Owner's response focuses on a subset of the claim elements in the independent claims that it contends are not disclosed by the reference cited by Petitioner. *Id.* Thus, Patent Owner's response to Petitioner's first ground of challenge focuses on claim elements 1g, 1i, 1h, and the corresponding elements in independent claims 16 and 20. *Id.* at 30–47. Patent Owner's response to petitioner's second anticipation ground focuses on claim elements 1i, 1d, 1e, and the corresponding elements in independent claims 16 and 20. *Id.* at 50–57. In addition, Patent Owner's response to both grounds focuses on claim elements 16b and 16j. *Id.* at 48–50.

⁵ While the Petition also mentions "Applicant Admitted Prior Art" (Pet. 10), Petitioner's anticipation analysis is based only on DaCosta or Kraft. Pet. 4, 14–69.

In our analysis of the merits, we focus on the disputed claim elements in the independent claims, particularly claim 1.⁶ As a result of our conclusions on these disputed elements, we do not need to reach the other claim elements or the dependent claims. Nor do we reach the issue of discretionary denial in view of our disposition of the merits. *See infra*, Section IV.

B. Anticipation

The Federal Circuit addressed the legal standard for anticipation in *Blue Calypso, LLC v. Groupon, Inc.*, 815 F.3d 1331 (Fed. Cir. 2016): “Under 35 U.S.C. § 102(b), a prior art reference will anticipate if it ‘disclose[s] each and every element of the claimed invention . . . arranged or combined in the same way as in the claim.’” 815 F.3d at 1341 (citation omitted) (footnote omitted).

C. Level of Ordinary Skill in the Art

According to Petitioner, a person of ordinary skill in the pertinent art “would have been an individual who, as of July 18, 2002, the filing date of the ’667 Patent, 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. 9 (citing Williams Decl. ¶¶ 30–31).

Patent Owner does not provide a description of the person of ordinary skill or dispute Petitioner’s description.

⁶ With the exception of claim limitations 16b and 16j, Petitioner relies on its analysis of claim 1 for independent claims 1, 16, and 20. *See* Pet. 38–41, 64–66 (claim 16), 42–44, 67–69 (claim 20),

We adopt the formulation of the level of ordinary skill in the art proposed by Petitioner as it is consistent with the prior art before us in this proceeding.

D. Claim Construction

Neither party proposes claim constructions for our consideration. Petitioner “gives the claim terms their plain and ordinary meaning in light of the specification and prosecution history as would have been read by a [person of ordinary skill].” Pet. 9. Patent Owner does not comment on claim construction.

We construe claim terms only as relevant to the parties’ contentions and only to the extent necessary to resolve the issues in dispute. *See Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999); *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017). We do not see the need to construe claim terms to resolve the issues in dispute.

E. Description of the Prior Art References

1. DaCosta (Ex. 1004)

DaCosta is a patent titled “System for Providing Database Functions for Multiple Internet Sources.” Ex. 1004 (54). DaCosta “relates to acquisition of data and, more particularly, to web browsers for the Internet, as well as to database utilities for data accessible through the Internet.” *Id.* at 1:22–29. According to DaCosta, “[t]he number of users professionally using the Internet (and particularly the ‘World Wide Web’) as a data source, and hence analogously to a database or collection of databases, on a daily basis is increasing.” *Id.* at 1:32–35. DaCosta explains that the web “merely brings up individual web pages to individual users. Unfortunately, these

web pages are typically depicted as HTML ‘pictures’ of data, and usually not the data itself.” *Id.* at 1:40–43.

DaCosta describes several distinct systems and embodiments. The preferred embodiment “performs the tasks of precisely locating and extracting the select data with a granularity specified by the user from any information source such as search engine results, web pages, other web-accessible documents, e-mail or text feeds in any format.” *Id.* at 2:31–36.

a) Navigation and Extraction Modules

The preferred system described by DaCosta comprises a navigation module that accesses one or more web pages or other web-accessible documents. *Id.* at 2:55–57. The navigation module provides the capability for a user to specify and store a procedure such as a series of clicks and entries of information, for example, a user name and password, to access a web page or other web-accessible document, as well as the capability to perform the procedure to actually access the web page or other web-accessible document in an automated manner. *Id.* at 2:57–64. The preferred system also comprises an extraction module that scrapes information from the accessed web page or other web-accessible document. *Id.* at 2:64–67.

Figures 1 and 7 of DaCosta are reproduced below, side-by-side.

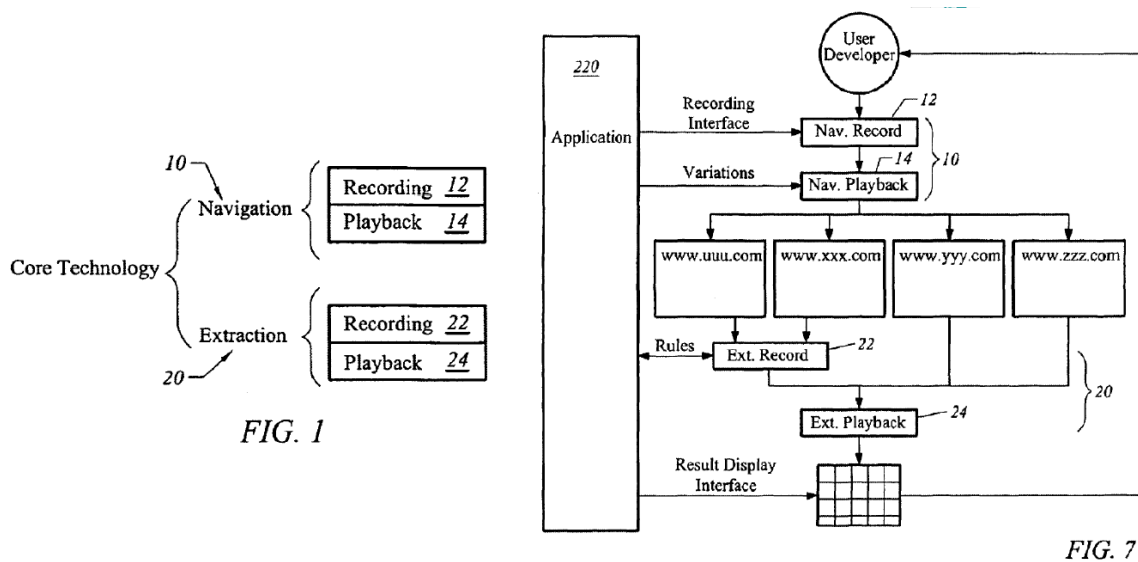


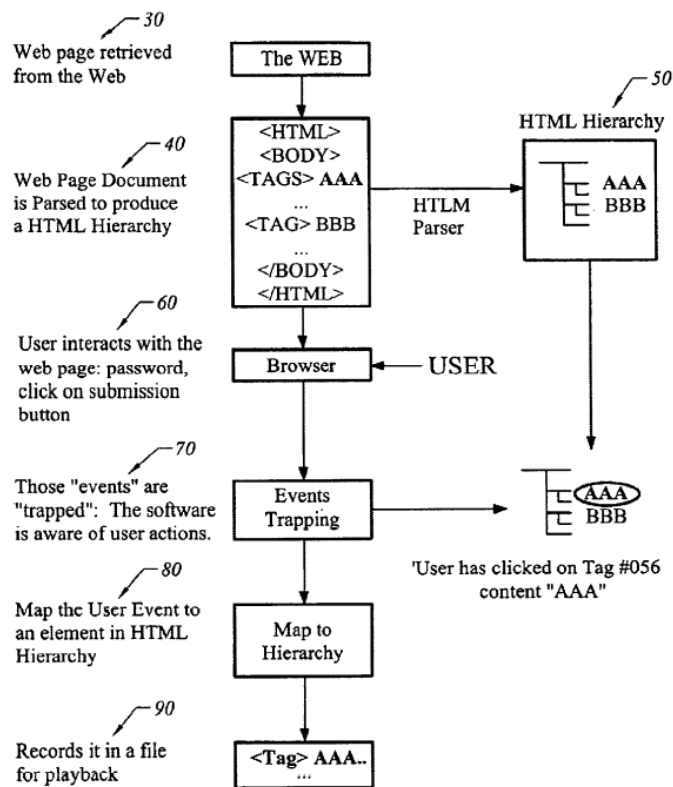
Figure 1 illustrates a core layer on top of which automated web applications may be built. *Id.* at 5:30–32. This provides a tool usable by developers or web users to access data available within web pages and other web-accessible documents, in an automated fashion. *Id.* at 5:32–36. Figure 7 illustrates an overview of a system integrating Navigation Application Program Interface (“API”) 10 and Extraction API 20 illustrated in Figure 1. *Id.* at 4:28–29, 10:55–57.

Navigation API 10 illustrated in Figure 1 enables a client application program running on a microprocessor-based device of a user to learn and store navigation paths to given web pages or other web-accessible documents, including dialogs and forms that need to be filled in to reach those locations or sites. *Id.* at 5:36–55. Navigation API 10 includes recording module 12 (which provides recording of navigation paths) and playback module 14 (which provides playback of that which has been recorded). *Id.* at 5:43–45, 6:51–59.

Extraction API 20 enables an application to define data segments in a web page or other web-accessible document, and includes recording module

22 (which provides recording of extraction patterns) and playback module 24 (which provides playback of extraction patterns). *Id.* at 5:55–6:31. Extraction module 20 records and plays across elements within a single page. *Id.*

Figure 2 of DaCosta, reproduced below, illustrates a flow diagram for recording a navigation script employing navigation recording module 12 of navigation API 10. *Id.* at 6:57–62.



Recording Navigation Script - Getting to a page

FIG. 2

Navigation recording module 12 records navigation paths and associated steps. *Id.* at 6:57–7:15. Each web page/web-accessible document (e.g., HTML web page) is provided (operation 30) to navigation module 10 by a

web application or a web browser such as Microsoft Internet Explorer (“IE”). *Id.* Each provided HTML web page/web-accessible document enters navigation module 10 and is converted (operation 40) into an element or object hierarchy. *Id.* Hierarchy objects or elements include lists, tables, images, and columns of the HTML page. *Id.* The object hierarchy is cached and used during the navigation recording operation (50). *Id.*

Once the web page or other web-accessible document has loaded, a user interacts (operation 60) with the loaded page by clicking on, or activating links or buttons, or entering data, and navigation recording module 12 captures (operation 70) each user-generated event (operation 60) such as clicks and keyboard inputs at the HTML level, using the built object hierarchy. *Id.* at 7:16–27.

Figure 4 of DaCosta, reproduced below, illustrates a flow diagram for recording an extraction script employing extraction API 20. *Id.* at 8:61–65.

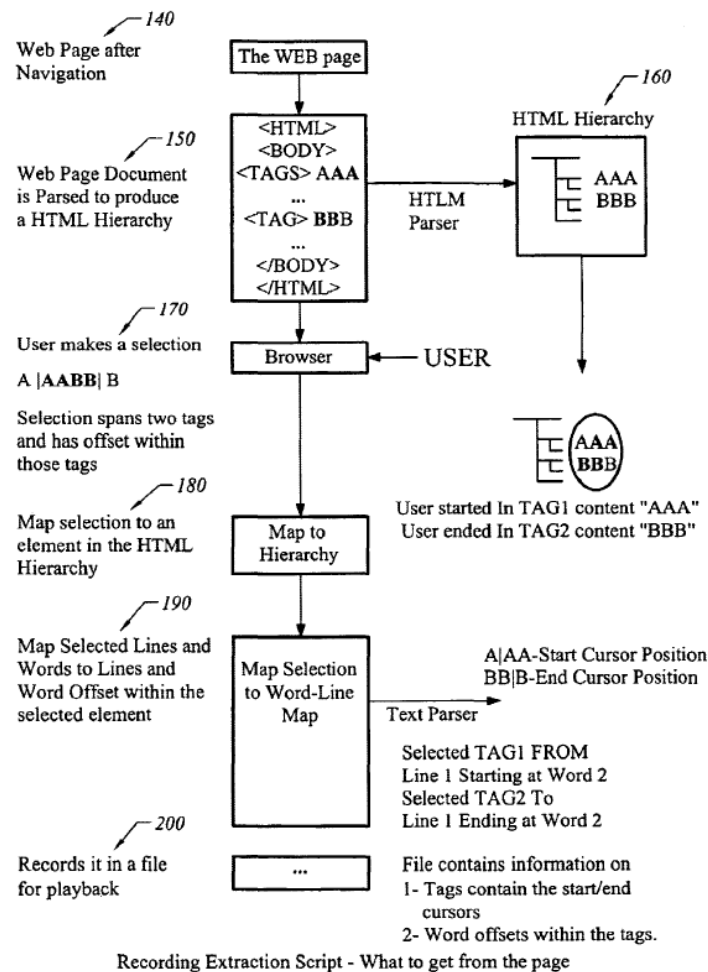


FIG. 4

While navigation module 10 operates across multiple web pages, extraction module 20 operates at the HTML or web page level. *Id.* at 8:61–9:4.

Extraction module 20 preferably follows navigation module 10 by processing each final or target web page or other web-accessible document loaded in response to operation of the navigation module, to extract relevant or select information or data therefrom. *Id.* Extraction recording module 22 accepts as input text selections within an HTML page, and a pattern is identified from these selections according to which data will be extracted using extraction playback module 24. *Id.* at 9:5–25.

More particularly, after an HTML web page/web-accessible document is loaded (operation 140), the page is parsed (operation 150) to produce the HTML hierarchy (operation 160). *Id.* at 9:52–10:5. The user then makes a selection from the web page/web-accessible document by highlighting a desired portion (operation 170)—for example, in Figure 4, the user has selected a starting element portion AA, part of parent element A, and ended the selection on element portion BB of element B. *Id.* That selection is then mapped (operation 180) into the HTML hierarchy that was created (operation 160). *Id.* The extraction script is then recorded (operation 200). *Id.* The output of extraction recording module 22 may be an XML file describing the selection(s) to be given to extraction playback module 24. *Id.* at 10:5–9.

b) Smart Clients

Navigation and extraction modules 10 and 20 may comprise plug-ins in a browser such as Microsoft IE, forming a smart client. *Id.* at 10:18–27. DaCosta explains that the power of smart clients lies in their capability to “meta-surf,” i.e., drive the browser to navigate to web sites and scrape relevant snippets of information from those sites. *Id.*

c) Smart Servers

An alternative “smart server” system is described by DaCosta. Ex. 1004, 18:8–32. The “smart server,” side-by-side with the separate “smart client” embodiment, is illustrated in Figure 22 as follows.

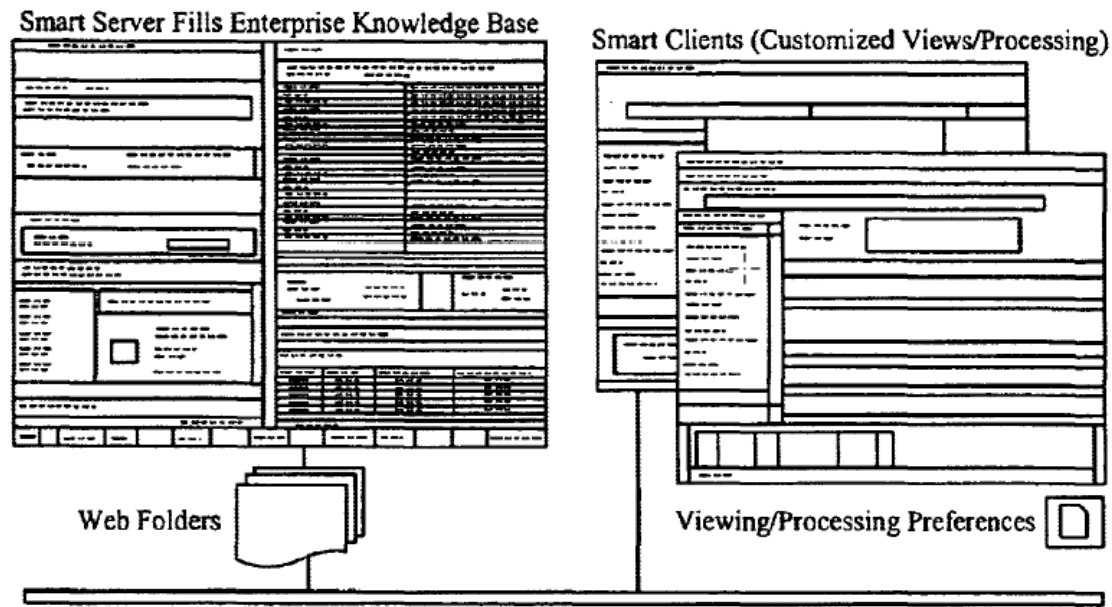


FIG. 22

Figure 22 shows the architecture for a smart web server for “scouring the internet” side-by-side with the alternative architecture for smart clients. Ex. 1004, 5:1–3.

In contrast to a “smart client,” a “smart server” meta-surfs to gather data on the Internet, with one of the roles of a smart server being “to fill a knowledge repository in background mode.” *Id.* at 18:8–26. For example, a smart server collects data required by stock analysts (e.g., recent news on various companies), providing distributed, multiple dynamic virtual data feeds. *Id.* DaCosta provides that “[t]he data feeds Web Folders that comprise a knowledge repository accessible by smart clients,” but in “an alternative embodiment . . . extracted data can be stored in a conventional repository, including Microsoft Exchange 2000 or any other database that constitutes a knowledge repository for XML tagged data” so that “the user can readily access the stored data needed in conjunction with other business processes.” *Id.* DaCosta further provides that “[e]ach request by the smart

server preferably ‘loops’ after a preset delay to visit the target web site and extract updated information.” *Id.*

d) Scripts

In a separate embodiment, DaCosta discloses that “scripts are stored at a centralized repository that is accessible through the Internet.” *Id.* at 18:33–35.

2. Kraft (Ex. 1005)

Kraft is titled “System and Method for Associating a Destination Document to a Source Document During a Save Process.” Ex. 1005, (54). Kraft relates to a software system and method “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.” *Id.* at ¶ 1. Kraft’s document association system saves the location of the source document, a query parameter, and other relevant input information as attributes of the saved document. *Id.* at (57). In addition, the system allows a user to return to the source document if a target or intermediary document is deleted. *Id.*

Figure 1 of Kraft follows.

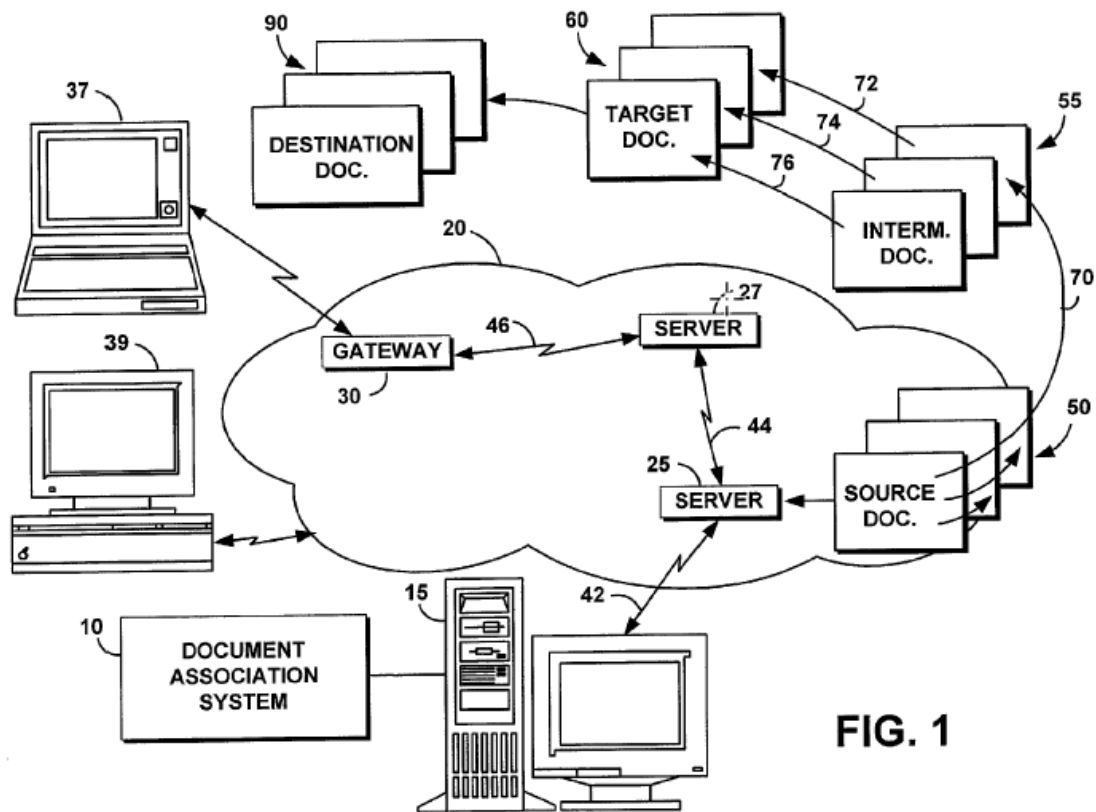


FIG. 1

Figure 1 is a schematic illustration of an exemplary operating environment in which a document association system described by Kraft is used.

Ex. 1005 ¶ 18. System 10 includes a software or computer program product that is typically embedded within, or installed, at least in part, on host server 15. *Id.* ¶ 40. Hypertext documents 50 (source document), 55 (intermediate document), 60 (target document) would include embedded hypertext links to other locally stored pages, and hypertext links 70, 72, 74, 76 to other webs sites or documents 55, 60 that are stored by various web servers such as server 27. *Id.* ¶ 42.

Figure 3 of Kraft is reproduced below.

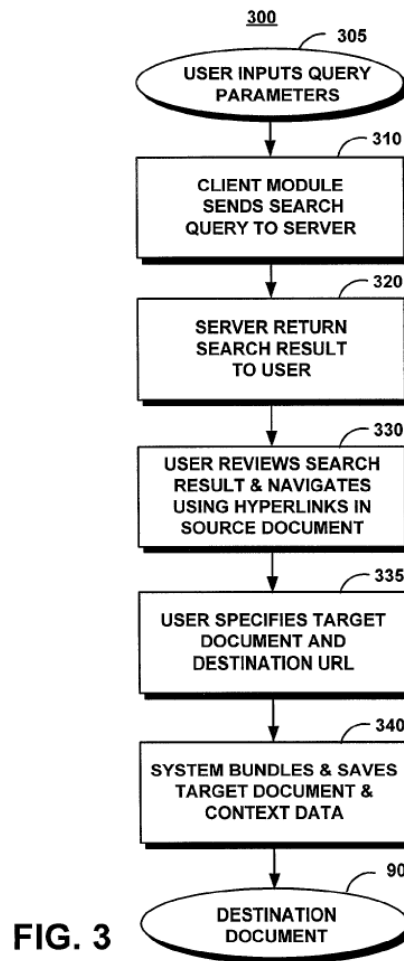


Figure 3 is a flowchart that illustrates specific steps of method 300 of operation of Kraft's document association system 10. *Id.* ¶ 20. Kraft's method starts with step 305, where the user inputs query parameters using a browser. *Id.* ¶ 57. At step 310, the document storage manager sends the search query to a service provider. *Id.* In step 320, the service provider returns search results to the user's web browser as the source document and establishes a connection with document association system 10. *Id.* ¶ 58.

The user reviews the search results (step 330) and navigates the Internet (step 335) using hyperlinks in the source document. *Id.* The user

continues the navigation until the user detects the desired target document, at which point the user identifies the target document, issues a save command, and enters the destination address (URL) of a repository where the user desires to store the destination document. *Id.* ¶ 59.

At step 340, a save command prompts the document storage manager to create the destination document by bundling the target document with the context data of the source document. *Id.* ¶ 60; *see also id.* ¶ 27 (a destination document is a “final document or web page which is comprised of a target document that is bundled with contextual data about the source document”). For example, when the source document is an HTML page including hyperlinks to various other documents, with one embedded hyperlink referencing a pdf target document titled “White Paper,” the user saves this target document to a storage medium by selecting the hyperlink; the target document “White Paper” is then bundled with context attributes, and is saved as a destination document that resides on the selected storage medium as a pdf document. *Id.* ¶¶ 61–62.

F. Anticipation of Claims 1–20 Based on DaCosta

1. Claim 1

Petitioner asserts that claim 1 was anticipated by DaCosta. Pet. 14–26. Petitioner provides an element-by-element claim analysis, supported by expert testimony. *See id.*; Williams Decl. ¶¶ 64–93.

Patent Owner disputes this analysis as failing to disclose “each of the features as recited in the claims.” Prelim. Resp. 1, 29–49. According to Patent Owner, 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).” *Id.* at 1. As discussed *infra*, in the following analysis we will focus on claim 1 and discuss only the disputed limitations.

a) *Limitation 1g*

Limitation 1g recites “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.” Pet. 20. In analyzing limitations 1d and 1e, which describe parsing and selecting the data elements, Petitioner relies on DaCosta’s disclosure of users selecting text from a web page as described in connection with the Figure 4 “extraction module” embodiment. Pet. 17–18; *see description supra*, Section III.E.1.a. For the recited storage of elements in a target data base, Petitioner cites the “smart web server” embodiment. Pet. 20 (citing Ex. 1001, 18:18–23); *see description supra*, Section III.E.1.c. Petitioner asserts that “DaCosta discloses ‘an architecture for a smart web server to scour data from the Internet for deposit in a data repository.’” Pet. 20.

For the claimed “selecting the target database,” Petitioner contends that DaCosta discloses selecting a particular target database and location “through the use of scripts.” *Id.* at 21. Petitioner continues, “through a navigation script 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” *Id.* (citing Ex. 1004, 5:12–20; Williams Decl. ¶ 81) (internal quotation marks omitted). Citing DaCosta’s description of another “preferred embodiment,” Petitioner also contends that scripts “may also be stored in a decentralized target

database.” Pet. 24 (citing Ex. 1004, 18:17–23; Williams Decl. ¶ 81); *see* description *supra*, Section III.E.1.d).

Patent Owner contends Petitioner’s analysis is “unclear.” Prelim. Resp. 31. Patent Owner challenges Petitioner’s reliance on the “smart server” embodiment. Prelim. Resp. 32–35. Patent Owner argues that “DaCosta’s smart server operates in the background mode to fill a knowledge repository with data. . . . But this data is different from the data selected from the web page via the extraction module 20.” *Id.* at 32 (citing Almeroth Decl. ¶¶ 59–60, 69). Patent Owner asserts also that the description in DaCosta of storing scripts in a central repository “is irrelevant to the user-selected text that is extracted from the webpage.” *Id.* at 35–36 (citing Almeroth Decl. ¶ 77). In support, Patent Owner compares the description of a user selecting data from a web page in the “extraction module” system of Figure 4 (Ex. 1004, 9:56–65) with the storage of scripts discussed in connection to a separate “scripts” embodiment (*id.* at 18:33–36).

We agree with Patent Owner that Petitioner has failed to demonstrate sufficiently that this limitation is disclosed by DaCosta as is required for showing anticipation. As discussed *supra*, “a prior art reference will anticipate if it ‘disclose[s] each and every element of the claimed invention . . . arranged or combined in the same way as in the claim.’” *Blue Calypso*, 815 F.3d at 1341. As noted, DaCosta discloses several distinct systems that operate independently, including the extraction module system of Figure 4, smart clients, and smart servers. *See* Ex. 1004, 16:10–13 (“Two preferred deployment vehicles for run time execution are provided, namely, smart

servers and smart clients”), Fig 22 (showing, separately, a smart client and a smart server).

DaCosta describes “smart servers” separately from “smart clients.” *Id.* at 18:8–33. According to DaCosta, “[s]mart servers provide distributed, multiple dynamic, virtual data feeds.” *Id.* at 18:14–15. Further, DaCosta describes that the data from smart servers feeds Web Folders that comprise a knowledge repository “accessible by smart clients” or alternatively, “can be stored in a conventional repository.” *Id.* at 18:15–19.

Focusing on smart servers as an example, DaCosta describes them as operating in “background mode” and as “scour[ing] data from the Internet for deposit in a data repository.” Ex. 1004, 5:1–3. Petitioner has not directed us to a disclosure showing that DaCosta’s smart servers store the “data elements” identified by Petitioner in its analysis of elements 1d and 1e that have been selected by a user from a web page using the extraction process depicted in Figure 4. Instead, DaCosta discloses that the smart servers operate in the “background.” DaCosta does not describe any connection between the smart server and the separately described system using navigation and extraction modules and the parsing of data into lists discussed elsewhere in DaCosta. Nor does Petitioner describe a connection between these separate disclosures. As Patent Owner asserts, we find that Petitioner identifies no disclosure in DaCosta of the smart server storing user-selected text as in the other systems described in DaCosta. Prelim. Resp. 35.

Nor has Petitioner provided us with an explanation showing that it would be proper to combine the various systems disclosed in DaCosta. The Federal Circuit has instructed us that “[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) (quoting *Connell v. Sears, Roebuck & Co.*, 722 F.2d 1542, 1548 (Fed.Cir.1983)). In *Net MoneyIN*, a reference relied on to show anticipation contained two separate “protocols,” neither of which contained all the elements “arranged or combined in the same way as claimed.” 545 F.3d at 1370. The Federal Circuit, reversing a finding of anticipation by the district court, concluded that “[t]he district court was . . . wrong to combine parts of the separate protocols shown in the . . . reference in concluding that [the challenged claim] was anticipated.” *Id.* at 1371. In reaching this conclusion, the Court distinguished between obviousness and anticipation: “Differences between the prior art reference and a claimed invention, however slight, invoke the question of obviousness, not anticipation.” *Id.* The Federal Circuit concluded that “it is not enough [for anticipation] 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.” *Id.*

For the reasons given, we find that Petitioner has failed to demonstrate that DaCosta anticipates claim 1. Petitioner has failed to demonstrate that these distinct systems disclosed by DaCosta are “‘arranged or combined in the same way as in the claim.’” *Blue Calypso*, 815 F.3d at 1341; *Net MoneyIN*, 545 F.3d at 1369–71. Petitioner relies on separate, distinct systems disclosed in DaCosta that are not linked, and in fact are separate in

their purpose and operation. The “smart server” embodiment operates in the background, and without user input “scours” the Internet for content to store in a repository. Ex. 1004, 5:1–3, 18:26–28. There is no user selection of the collected data described. In contrast, the “smart client” embodiment using navigation and extraction scripts allowing the user to select the data elements to be extracted displayed. *Id.* at 11:12–19.

For the reasons given, we find that Petitioner fails to demonstrate sufficiently that DaCosta teaches this claim limitation.

b) Limitation 1h

Limitation 1h recites “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.” Pet. 22. Petitioner contends DaCosta discloses this step. *Id.* at 22–24. Petitioner contends that “DaCosta discloses the use of scripts, also referred to as Navex scripts, that include source and destination information, e.g., a location file.” *Id.* at 22–23 (citing Ex. 1004, 11:19–24; Williams Decl. ¶¶ 83–84). Petitioner further contends, “DaCosta also discloses saving the target location in a database.” *Id.* at 23 (citing Ex. 1004, Fig. 10; Williams Decl. ¶ 85). Petitioner identifies the “target data base” as “a conventional repository or centralized repository.” *Id.* (citing Ex. 1004, 18:17–23, 18:34–38, Fig. 10; Williams Decl. ¶¶ 86–87).

Patent Owner responds that Petitioner has failed to meet its burden because Petitioner identifies “the URL or metabookmark as corresponding to the source location” and “identifies two distinct repositories—the knowledge repository for *data* and centralized repository for *scripts*—as corresponding to the target location and target database.” Prelim. Resp. 43. Patent Owner argues further that DaCosta’s navigation and extraction scripts (the Navex

scripts) do not disclose a target location in a target database. *Id.* at 44–46. Patent Owner further argues that Petitioner’s analysis of the databases is “incompatible with DaCosta’s teachings.” *Id.* at 46–47.

We agree with Patent Owner that Petitioner’s analysis of this claim limitation “misconstrues the functions of the two separate databases to create something that is not disclosed in DaCosta.” *Id.* at 47 (citing Almeroth Decl. ¶ 90). Dr. Almeroth testifies that “[t]he conventional repository is an alternative embodiment for storing *extracted data* from the smart server and the centralized repository is used for storing scripts.” Almeroth Decl. ¶ 90 (citing Ex. 1004, 18:8–23, 18:34–40). He testifies further that “DaCosta discloses a first database for storing scripts (the centralized repository) and a second, distinct database for storing data extracted by the smart servers (the knowledge repository).” *Id.* Further, “[a person of ordinary skill] would not understand that scripts (or other information) would be stored in the knowledge repository, or that extracted data would be stored in the centralized repository.” *Id.*

We find these arguments by Patent Owner and the supporting testimony to be supported by the disclosures of DaCosta and convincing. As it has previously, Petitioner again relies on separate systems disclosed in DaCosta that are not linked, and in fact are separate in their purpose and operation. For the reasons given, we find that Petitioner fails to demonstrate sufficiently that DaCosta teaches this claim limitation.

c) Limitation 1i

Limitation 1i recites “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.” Pet. 24.

Petitioner contends DaCosta discloses this step. *Id.* at 24–25. Petitioner contends that “DaCosta discloses repeating its method of extracting, processing, [and] integrating.” *Id.* at 24–25 (citing Ex. 1004, 5:12–20, 7:62–65, 18:23–25).

Patent Owner asserts that Petitioner has failed to meet its burden for this claim limitation. Prelim. Resp. 40–43. Patent Owner argues that “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.” *Id.* at 40. Patent Owner also contends that “none of the steps identified by [P]etitioner disclose transferring data to a database.” *Id.* at 41. Patent Owner gives, as one example, Petitioner’s citation to the smart server “looping ‘after a *preset* delay.’” Prelim. Resp. 42–43.

We agree that the examples in DaCosta cited by Petitioner do not disclose setting a schedule frequency as recited in this limitation. As Dr. Almeroth testifies, responding to Petitioner’s “generic” assertion of repetition by its disclosed method, “neither ‘extracting,’ ‘processing,’ nor ‘integrating’ disclose transferring the parsed data element to the target database.” Almeroth Decl. ¶ 81. Responding to Petitioner’s argument based on “looping” by the smart server, Dr. Almeroth testifies, “the smart server is a distinct embodiment that stores data in a background mode and is unrelated to the user oriented processes that [Petitioner’s expert] Dr. Williams maps to as corresponding to the parsed data elements.” Almeroth Decl. ¶ 84.

We find these arguments and the supporting testimony by Dr. Almeroth to be persuasive. Among other reasons, we find that the user-

selected text that Petitioner identifies as the claimed “at least one of the data elements” is not stored in the knowledge repository or utilized by the smart server. Thus, the operation of the smart server cited by Petitioner does not disclose claim limitation 1i. Prelim. Resp. 40–43. Nor are we persuaded that DaCosta’s disclosure of automatically repeating certain operations teaches setting a schedule for transferring data for the reasons given by Patent Owner. *See* Pet. 24–25; Prelim. Resp. 41–42.

Accordingly, for the reasons given, we find that Petitioner fails to demonstrate sufficiently that DaCosta teaches this claim limitation.

d) Conclusion – Claim 1

For the reasons given, we are persuaded that Petitioner has not demonstrated that at least imitations 1g, 1h, and 1i are sufficiently disclosed by DaCosta. We, therefore, determine that the Petition does not demonstrate a reasonable likelihood of prevailing on this anticipation challenge to claim 1.

2. Independent Claims 16 and 20

Petitioner presents contentions that independent claims 16 and 20 were anticipated by DaCosta. Pet. 38–44. Petitioner relies on its analysis of claim 1 for these claims. *See supra*, Section III.A. Patent Owner opposes based on its arguments regarding claim 1 limitations 1g, 1i, and 1h. Prelim. Resp. 29–47. For the reasons given *supra* with respect to the limitations 1g, 1h, and 1i in connection with claim 1, we find that Petitioner fails to demonstrate sufficiently that DaCosta teaches the corresponding imitations in claims 16 and 20.

With respect to claim 16, Patent Owner additionally argues that Petitioner has failed to meet its burden for limitations 16b and 16j. Prelim.

Resp. 48–49. Those limitations recite, respectively, “a first set of instruction codes for initializing a data retrieval stage” and “a second set of instruction codes for 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, second sets of instruction codes.” Pet. 9, 41.

To meet these limitations, Petitioner relies on the disclosed navigation and extraction scripts in DaCosta, respectively, as the first and second set of instructions. Pet. 39, 41. Petitioner refers back and incorporates by reference its discussion of claim elements 1b (describing initializing) and 1c (describing parsing). *Id.* at 39, 41.

Patent Owner argues that Petitioner’s approach is “inconsistent” because “Petitioner relies on the extraction script—the alleged second set of instruction codes—for teaching several different elements of Limitations [1c]-[1i].” Prelim. Resp. 48; Almeroth Decl. ¶¶ 92–93. Petitioner adds that “Petitioner analysis of the second set of instruction codes also includes similar inconsistencies.” *Id.* at 49.

In view of the foregoing determination that other claim limitations are not met by DaCosta, we do not reach these arguments addressing additional limitations of claim 16.

3. *Dependent Claims 2–15 and 17–19*

Claims 2–15 and 17–19 depend either from independent claim 1 or 16. We find that Petitioner’s anticipation challenges to dependent claims 2–15 and 17–19 based on DaCosta are insufficient for the reasons explained above for these independent claims.

Accordingly, for the reasons given for claims 1 and 16, we are not persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing on this anticipation challenge as to dependent claims 2–15 and 17–19.

4. *Conclusion*

In sum, after considering the arguments presented in the Petition and Patent Owner’s responses, we determine that, for the reasons given, the Petition does not demonstrate a reasonable likelihood of prevailing on these anticipation challenges based on DaCosta.

G. *Anticipation of Claims 1–11 and 13–20 Based on Kraft*

1. *Claim 1*

Petitioner asserts that claim 1 was anticipated by Kraft. Pet. 44–56. Petitioner provides an element-by-element claim analysis, supported by expert testimony. *See id.*; Williams Decl. ¶¶ 159–183.

Patent Owner opposes. Prelim. Resp. 49–57. Patent Owner’s opposition focuses on claim limitations 1d, 1e, and 1i. Our analysis will address claim limitation 1e.

a) *Limitation 1e*

Petitioner contends Kraft discloses limitation 1e (“selecting at least one of the data elements from the list to be stored in the target database”). Pet. 49. Petitioner contends that “Kraft discloses that ‘[t]he user continues his or her navigation until he or she detects the desired target document 60 (FIG. 1).’” *Id.* (citing Ex. 1005 ¶ 59). Petitioner continues that, at this point, “the user [in Kraft] identifies such target document 60, 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 (FIG. 1) and [t]he destination documents repository 170 can be located on the user's computer, on the network 20, and/or within the service provider 100.” (*Id.*, quoting Ex. 1005 ¶ 59) (internal quotation marks omitted) (alterations in original).

Patent Owner responds that Petitioner offers no explanation of how Kraft's disclosure teaches limitation 1e. Prelim. Resp. 53. Dr. Almeroth testifies that “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.” Almeroth Decl. ¶ 102 (emphasis added). He continues that “[t]here is no teaching [in Kraft] 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.’” *Id.*

We agree with this analysis by Dr. Almeroth as it is consistent with Kraft. Paragraph 59 of Kraft, cited by Petitioner, states that the user identifies *target document 60* and issues a save command and destination URL that results in saving the document to the designated repository. Ex. 1005 ¶ 59. In contrast, the claim calls for selecting data elements to be stored in a database from a list that results from parsing selected data elements. *See* claim limitation 1e in Section II.D, *supra*. We therefore agree with Patent Owner and find that because “Kraft discloses navigating to a target document 60 (e.g., a web page) and then saving the entire document,” Kraft does not disclose this limitation. Prelim. Resp. 53 (citing Almeroth Decl. ¶ 102).

As discussed *supra*, in Section III.E.2, a user's save command in Kraft prompts the document storage manager to create a destination document by bundling the target document with the context data of the

source document. Ex. 1005 ¶ 60; *see also id.* ¶ 27 (a destination document is a “final document or web page which is comprised of a target document that is bundled with contextual data about the source document”). Therefore, we find that Kraft does not teach selecting data elements from a list to be stored in a target database as claimed in limitation 1e. Accordingly, we determine that Petitioner fails to demonstrate sufficiently that Kraft teaches this claim limitation.

For the reasons given, we are persuaded that Petitioner has not demonstrated sufficiently that at least limitation 1e is disclosed by Kraft. We, therefore, determine that the Petition does not demonstrate a reasonable likelihood of prevailing on this anticipation challenge to claim 1.

b) Limitations 1d and 1i

Patent Owner challenges Petitioner’s analysis of these claim elements in light of Kraft. *See* Prelim. Resp. 50–53, 54–57. In view of the foregoing determination for limitation 1e, we do not reach these additional arguments.

2. Independent Claims 16 and 20

Petitioner presents contentions that independent claims 16 and 20 were anticipated by Kraft. Pet. 64–69. Patent Owner opposes based on arguments similar to the arguments presented with respect to limitation in 1e claim 1, as well as limitations 1d and 1i. Prelim. Resp. 50–57.

For at least the reasons given for limitation 1e in Section III.G.1.a, *supra*, we are not persuaded that Petitioner has demonstrated sufficiently that the corresponding limitation in claim 16 is disclosed by Kraft. We, therefore, determine that the Petition does not demonstrate a reasonable likelihood of prevailing on this anticipation challenge to claim 16. We apply the same reasoning to claim 20 and determine that the Petition does not

demonstrate a reasonable likelihood of prevailing on this anticipation challenge to that claim.

With respect to claim 16, Patent Owner additionally argues that Petitioner has failed to meet its burden for limitations 16b and 16j. Prelim. Resp. 50. In view of the foregoing determinations, we do not reach these additional arguments.

Accordingly, for the reasons given, we are not persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing on this anticipation challenge to independent claims 16 and 20.

3. Dependent Claims 2–11, 13–15, and 17–19

Claims 2–11, 13–15, and 17–19 depend from either independent claim 1 or 16. Petitioner’s anticipation challenges to dependent claims 2–11, 13–15, and 17–19 based on Kraft are insufficient for the reasons explained above for those independent claims.

Accordingly, for the reasons given for claims 1 and 16, we are not persuaded that Petitioner has demonstrated a reasonable likelihood of prevailing on this anticipation challenge as to dependent claims 2–11, 13–15, and 17–19.

4. Conclusion

In sum, after considering the arguments presented in the Petition and Patent Owner’s responses, we determine that, for the reasons given, the Petition does not demonstrate a reasonable likelihood of prevailing on these anticipation challenges based on Kraft.

IV. 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. 17–29. We do not reach those issues in light of our assessment of the merits of the Petition, as discussed above.

V. CONCLUSION

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

VI. ORDER

Upon consideration of the record before us, it is:

ORDERED that the Petition is *denied*, and no trial is instituted.

IPR2023-01442
Patent 7,020,667 B2

For PETITIONER:

Thomas M. Dunham
Robert W. Molitors
Elizabeth A. O'Brien
CHERIAN LLP
tom@dunham.cc
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