

I. The '796 Patent

~~A. Effective Filing Date~~

~~The '796 patent issued from Application No. 09/898,289, filed July 3, 2001. During prosecution, the applicants alleged “conception of an invention falling within independent claims 1, 12, 23-25 and at least one or more dependent claims, at least as early as March 14, 2001.” EX1002, 116. Petitioner does not concede that the applicants demonstrated conception as of March 14, 2001, but for purposes of this Petition only, Petitioner has used this date as the relevant priority date because all prior art cited herein was available before then.~~

~~B. Level of Ordinary Skill in the Art~~

~~A POSITA in the field of the '796 patent would have had a bachelor's degree in computer science, computer engineering or an equivalent, as well as four years of industry experience, and would have had working knowledge of data extraction methods and programming languages including perl and flex. EX1003 ¶23. Additional graduate education could substitute for professional experience, or significant experience in the field could substitute for formal education. Id.~~

~~C.A. Overview of the Challenged '796 Patent~~

~~5. The '796 patent relates to “information extraction.” from data streams such as text and spoken documents.” EX1001, ¶7-10. The~~Notably, the specification's “Background of the Invention” admits that ~~information~~previous data-extraction techniques ~~in the prior art~~ could “identify all occurrences of certain classes of words in a document,” but alleges that there remained “a need for data processing techniques which explicitly identify portions of data that are sought to be identified.”, rather than only implicitly identifying, or tagging, such portions of data.” *Id.*, 1:24-27, 1:46-49.

The '796 ~~patent~~ purports to solve this alleged problem by “using stereotypical phrases or ‘regular expressions’ to identify the information-bearing portions of a document.” ~~*Id.*, 2:65-3:2.~~

~~Regular, and thereby explicitly identity the information found therein.” *Id.*, 2:65-3:2. The patent states that regular~~ expressions “comprise symbolic strings, typically word sequences, that can be constructed by the basic operations of ‘and’ and ‘or’ acting on an atomic alphabet of symbols.” *Id.*, 3:2-8. For example, a message could include “the phrase ‘call me back at 555-1234.’” *Id.*, 11:27-29.

The alleged invention

6. _____ “determines that the phrase ‘call me back at’ matches one of the enumerated expressions,” and thus “the information associated with the phrase ‘call me back at,’ namely, ‘555-1234,’ is identified as the information-bearing portion of the expression, and extracted.” *Id.*, 11:29-34.

7. _____ The ’796 patent admits that regular expressions were known in the prior art and “have been discussed in Aho et al., ‘Compilers: Principles, Techniques, and Tools,’ Addison-Wesley, 1985.” *Id.*, 3:2-4. This was the standard textbook on compilers in the mid-1980’s and was my textbook when I first learned about regular languages and regular expressions around the same time. This book is also where I first learned the programming language “lex,” the forebear of “flex” (which simply stands for “fast lex”).

8. _____ The ’796 patent further admits that regular expressions may “be stored in one of several common programming languages, e.g., as a ‘flex’ or ‘perl’ program,” ~~both of which “have built-in support for regular expressions” (*id.*, 2:23-27) and were known in the prior art (*id.*, 3:26-31).~~

program,” both of which “have built-in support for regular expressions” (*id.*, 2:23-27) and were described in prior-art references published in 1992 and 1996, respectively (*id.*, 3:26-31). A POSITA as of March 2001 would have been closely familiar with these programming languages and with the use of regular expressions.

9. The '796 patent has five independent claims. Claim 1 recites:

1. A method of automatically processing an input sequence of data symbols, the method comprising the steps of:
 - identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression;
 - identifying at least a portion of information associated with the at least one regularly identifiable expression; and
 - extracting the portion of information.

10. Claims 12 and 23 restate the same alleged invention as an “[a]pparatus” or “article of manufacture,” respectively. Claim 24 recites the same apparatus as claim 12 but adds a generic “data capture device for obtaining the input sequence of data symbols.” Claim 25 recites the same method as claim 1 without the final extraction step and adding that regular expressions are identified “by comparing the input document to one or more previously-stored regular

expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions.”

11. The As discussed in greater detail below, the dependent claims generally recite this same step of “comparing” data to “match” (claims 2, 13); “normalizing” data (claims 3, 14); “classes of data symbols” (claims 4, 15); “a characteristic phrase” that precedes or follows information (claims 5, 6, 16, 17); taking any “specified action” (claims 7, 18); “present[ing]” information to a user (claims 8, 19); using the flex, lex, or perl programming languages (claims 9, 20); limiting data to categories including text, audio, or video (claims 10, 21); and limiting data to voice mail (claims 11, 22).

D.B. Relevant Prosecution History

12. I have reviewed and considered the prosecution history of the '796 patent, which is marked EX1002. I provide a brief summary of certain key events during the prosecution of the '796 patent below.

The Examiner initially rejected all claims as anticipated or obvious over Adler (U.S. Pat. No. 6,342,901). EX1002, 86-90. In response, the applicants added the clause, “wherein [the at least one regularly identifiable expression / a regular expression] represents a pattern that is matchable in accordance with a programming language

13. that supports such a regularly identifiable expression,” to all independent claims. *Id.*, 96-100. The applicants argued that Adler lacks this

limitation because it only teaches the use of “cue words,” which “do not represent patterns that are matchable in accordance with a programming language that ~~in accordance with a programming language that~~ supports such patterns, as in the claimed invention.” *Id.*, 101.

14. The Examiner then rejected the claims over Klein (U.S. Pat. App. Pub. No. US 2002/0194026), which teaches “[a] PERL language regular expression parser

... along with a rule set that includes a library of regular expressions that define how information is delimited.” *Id.*, 108-09. ~~The~~In response to the Examiner’s rejection, the applicants did not dispute that Klein reads on the claims. Instead, they submitted a declaration under 37 C.F.R.

§1.131 that allegedly shows ~~prior~~ “conception “of an invention falling within independent claims 1, 12, 23-25 and at least one or more dependent claims, at least as early as March 14, 2001, and thus prior to the June 13, 2001 effective date of Klein.” *Id.*, 116. The Examiner then allowed the claims: following this declaration. *Id.*, 139.

15. The prior art cited in this Declaration was not considered during prosecution. As ~~shown~~discussed below, it is my opinion that the prior art in this ~~Petition, which was not previously before the Office,~~Declaration teaches all limitations of the ’796 patent’s claims, including the “wherein” clause that was added during prosecution to distinguish Adler.

II. Claim Construction

16. For purposes Counsel for Petitioner has informed me that language in a patent claim should be read in accordance with the ordinary and customary meaning of this Petition

such claim as understood by a POSITA in light of the specification and prosecution history of the patent.

17. Counsel for Petitioner has also informed me that only, the those terms need be construed that are in controversy, and only claim construction that is warranted¹ is the correction of a typographical error in claim 20: to the extent necessary to resolve the controversy.

18. In my opinion, the reference to the “pen programming language” in claim 20 of the ’796 patent is an obvious typographical error that should be construed as “perl programming language.”

19. Courts may correct “I have been informed that obvious minor typographical and clerical errors in ~~patents~~ patent claims may be corrected so long as “(1) the correction is not subject to reasonable debate based on consideration of the claim language and the specification; and (2) the prosecution history does not suggest a different interpretation of the claims.” *Novo Indus., L.P. v. Micro Molds Corp.*, 350 F.3d 1348, 1357, 1354 (Fed. Cir. 2003). “Board decisions have adopted Novo’s two-part test In this case, the reference to correct obvious errors.” *Samsung Elecs. Co. v. M&K Holdings Inc.*, IPR2018-00696, Paper 55 at 13-14 (PTAB Aug. 23, 2021) (collecting cases).

Here, “pen” in claim 20 is clearly a typographical error that should read “perl.” EX1003

20. ¶45. “Pen Notably, “pen” is not referenced in any other claim ~~or of~~

the '796, the specification, and/or the prosecution history.

The prosecution history of the '796 patent confirms that “pen” is a typographical error in the issued claims because the last- submitted version of claim 20 before allowance correctly recites “perl.” EX1002,

21. 99. The visual similarity between the letter “n” in “pen” and the contiguous letters “rl” in “perl” suggests

that the inclusion of “pen” in the issued version of the ’796 patent may have resulted from an error in optical character recognition (“OCR”).

Moreover, dependent claims 13-22 (which depend from claim 12) recite

~~¹ This is not a concession as to the scope of any claim term in litigation or a waiver of any argument that claim terms are indefinite, invalid, or unpatentable.~~

22. materially identical limitations as claims 2-11 (which depend from claim 1), and claim 9—the counterpart to claim 20—correctly recites “one of the flex, lex, and perl programming language.” This mirrors the specification’s correct recitation of these same three programming languages. EX1001, 3:22-25. Thus, ~~correcting “pen”~~ (“In accordance with an illustrative embodiment of the invention, the term ‘regular expression’ is taken to “perl” is not subject to reasonable debate, and nothing mean any form of pattern that is matchable in the prosecution history suggests a different interpretation. EX1003 ¶48: flex, lex, or perl programming languages....”).

III.—Detailed Discussion of the Grounds for Unpatentability

23. ~~Overview~~ Thus, in my opinion, correcting “pen” to “perl” would not be subject to reasonable debate to a POSITA, and nothing in the prosecution history suggests a different interpretation.

24. For purposes of my opinions of unpatentability in this Declaration, no other claim construction is necessary because, in my opinion, the challenged claims would have been obvious over the cited references under any reasonable construction and based on their plain and ordinary meaning to a POSITA.

A.III. Summary of Cited Prior Art References

25. The prior art references relevant to my analysis are briefly summarized below. I discuss additional elements of each reference in my analysis of how a POSITA would have combined them to arrive at the challenged claims.

1.A. Sheth (EX1004)

I understand that Sheth was prior art to the '796 patent at least because it issued from an application filed on August 21, 2000, and is prior art under §102(e).

26. EX1004.

Sheth states that it “is directed to software, a system and a method for creating a database of metadata (metabase) of a variety of digital media content,” particularly Internet media content, on the Internet. *Id.*, 4:54-59.

27. “~~It~~ Sheth explains that, “[i]n order to keep the metabase up-to-date with the latest news or movie trailers, metadata extraction from various sources” is provided. *Id.*, 9:66-10:1.

Sheth describes an “extraction process during which metadata for digital media is discovered, copied from a Web site, enhanced, and inserted into the present invention’s database.” *Id.*, 10:7-9. This Sheth’s extraction process utilizes “WebCrawlers 3 that scan certain Web sites” and “Extractors 7 that do the actual metadata extraction.” *Id.*, 10:10-12. Sheth explains: “Crawling and extraction rules are input to WebCrawler and Extractor

28. programs that traverse Web sites and retrieve digital media metadata

from selected pages.” *Id.*, 10:40-42.

Once

```
# Name : CNNUS
#
producer           [c]CNN
description         [] (?i)<img src[^>]*alt="Video"[^>]*>([<]*)<
descriptionReplace1 s/\([^\(\)]*\)\//
parentURL          [u](.*)
invalidated        [c]0
surrogate          [] (?i)([<]*)</h.>
productionSource   [c]tv
location           [S] (?i)/date.*?<p>(.*)\\(.*?\\)
locationReplace1   s/. *?<I>.*?<\\I>(.)/$1/
locationReplace2   s/(.*)--.*?/$1/
locationReplace3   s/. *?<em>.*?<\\em>(.)/$1/
mediaType          [c]video
source             [c]http://www.cnn.com
```

29. In Sheth, once a webpage is identified for extraction, “[t]he extractor scans the Web page content for pieces of text that match the pattern specified by the extraction rules.” *Id.*, 11:50-52. “After Furthermore, “[a]fter the extractor has attempted to find every attribute in its list of extraction rules, it creates an XML document containing attribute-value pairs.” *Id.*, 11:52-54, Fig. 6. Sheth explains that “[t]he ability to retrieve text that matches a particular pattern is provided by the PERL regular expression language.” *Id.*, 11:64-66; *see also id.*, 6:45-47. Examples of extraction rules written in perl are provided in Figures 8 and 9. *Id.*, 11:60-64, 12:23-33, Figs. 8-9; EX1003 ¶¶54-8, 9. Figure 8 is reproduced below:

```
# Name : CNNUS
#
producer           [c]CNN
description         [ ](?i)<img src[^>]*alt="Video"[^>]*>{[^<]*}<
descriptionReplac1 s/\[([^\]]*)*\]\[\\//
parentURL          [u](.*)
invalidated        [c]0
surrogate          [ ](?i){[^<]*}</h.>
productionSource   [c]tv
location           [S](?i)/date.*?<p>(.*)\\(.*?\\)
locationReplac1    s/. *?<I>.*?<\\//I>(.*)/$1/
locationReplac2    s/(.*?)--.*?/$1/
locationReplac3    s/. *?<em>.*?<\\//em>(.*)/$1/
mediaType          [c]video
source             [c]http://www.cnn.com
```

EX1004, *Id.*, Fig. 8.

2.B. Richardson (EX1005)

30. I understand that Richardson was prior art to the '796 patent at least because it issued from an application filed in 1998 and is prior art under §102(e). EX1005.

Richardson states that it “is directed to a method and apparatus for automatically recognizing and responding to intent in message text, particularly as used in a large-scale automatic message response system to examine input text and respond appropriately with output text relevant to the writer’s intent.” *Id.*, 1:21-26.

31. To determine the writer’s intent, Richardson’s ~~invention~~ “system isolates the writer’s critical words which relate to prototypical statements stored in the system.” *Id.*, 2:56-57. ~~The~~ Richard explains that “[t]he input message text can be electronic mail and may include text” ~~or,~~ among other computer-readable data. *Id.*, 3:19-20.

32. Figure 1 depicts apparatuses for implementing Richardson’s ~~invention~~ method and system, whereby “[c]ustomer system 12 sends input message 14 to server 16 for processing.” *Id.*, 4:40-41, Fig. 1. “Server 16 is adapted to take an action based on an internal knowledge base and the text of query message 14,” and “[i]f the action is a response message 18, that response message 18 is sent to customer system 12.” *Id.*, 4:44-47. Customer system 12 and server 16 “are general purpose computer systems programmed according to their use and

comprising one or more central processing unit” and “memory,” among other components. *Id.*, 5:10-19.

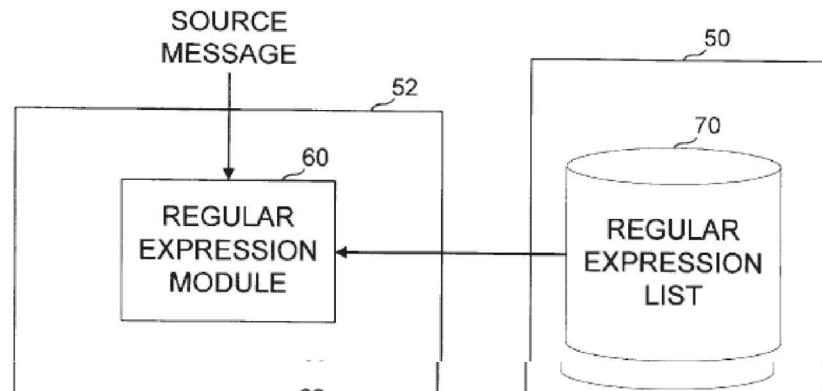
Server

Richardson's server 16 comprises, among other components, "knowledge base 36" and "message processing module

33. 38." *Id.*, 4:60-61, 5:4-6,

5:23-24, Figs. 2-3. In turn, knowledge base 36 comprises

“extraction rule set 50,” and message processing module 38 comprises “natural language extraction (NLE) module 52,” which are partly depicted below in Figure 4. *Id.*, 5:25-29.



Id., Fig. 4 (detail).

34. As Richardson explains, “FIG. 4 shows NLE module 52 comprising a regular expression module 60,” and “[e]xtraction rule set 50 is shown comprising a regular expression lexicon 70.” among others. *Id.*, 5:29-34. Understanding In Richardson, “understanding and responding to text messages begins with accepting the input text message, which is first processed by NLE module 52,” whereupon “[r]egular expression module 60 identifies instances of the regular expressions represented in regular expression lexicon 70.” *Id.*, 5:34-37, 5:51-57-59. The message and its extracted data are then sent to an intent matching module (IMM). *Id.*, 6:34-36.

35. Richardson provides a more detailed embodiment in Figs. 5(a)-(c) and 5(b), which depict how an “input message is processed by NLE module 52 to generate tags associated with portions of the input message.” *Id.*, 8:42-45. “NLE module 52 uses a regular expression list 104 and a phrase types list 106 in assigning tags to a message,” and “[a]n example of regular expression list 104 ... is presented in Table 1,” reproduced below. *Id.*, 8:50-55.

TABLE 1

Regular Expression List	
Expression	Type
http://*{ /* }>	URL
###-###-####	Phone
[a-zA-Z]@*,*	Email
...	...

message,” and “[a]n example of regular expression list 104 ... is presented in Table 1” below. *Id.*, 8:50-55.

TABLE 1

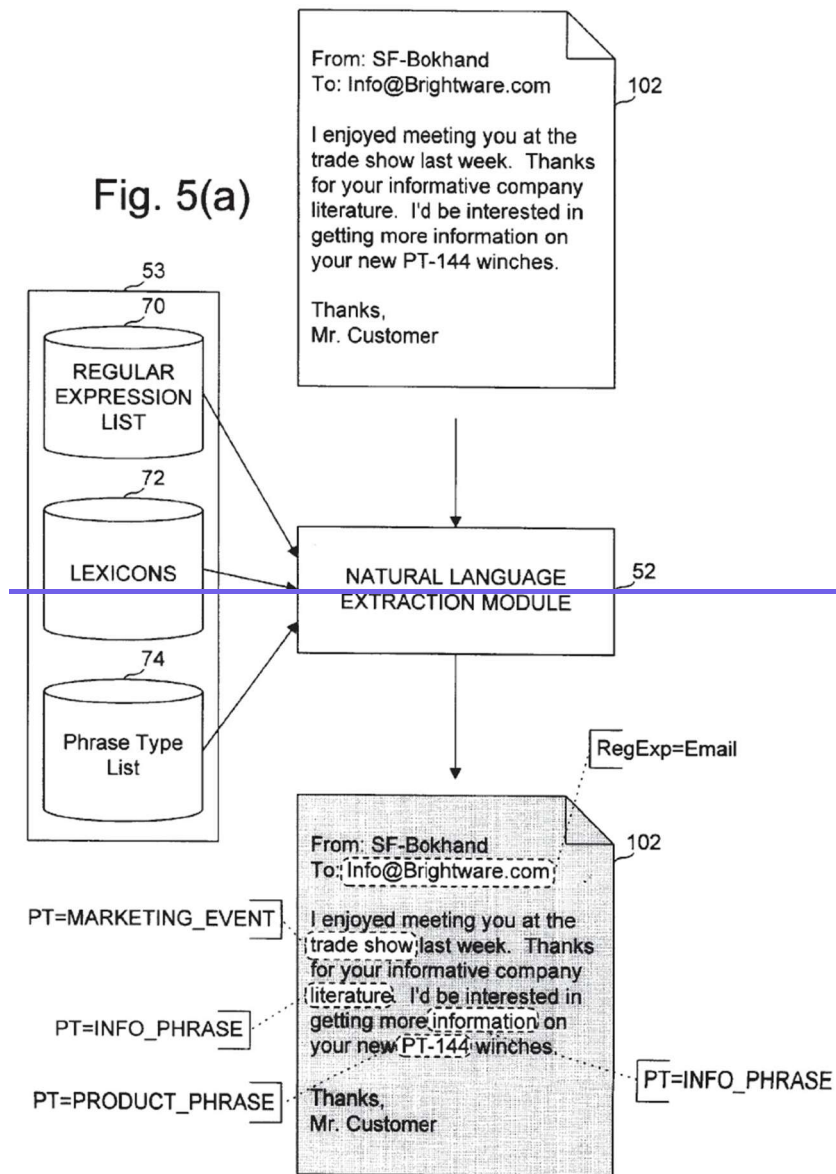
Regular Expression List	
Expression	Type
http://*{ /* }>	URL
###-###-####	Phone
[a-zA-Z]@*,*	Email
...	...

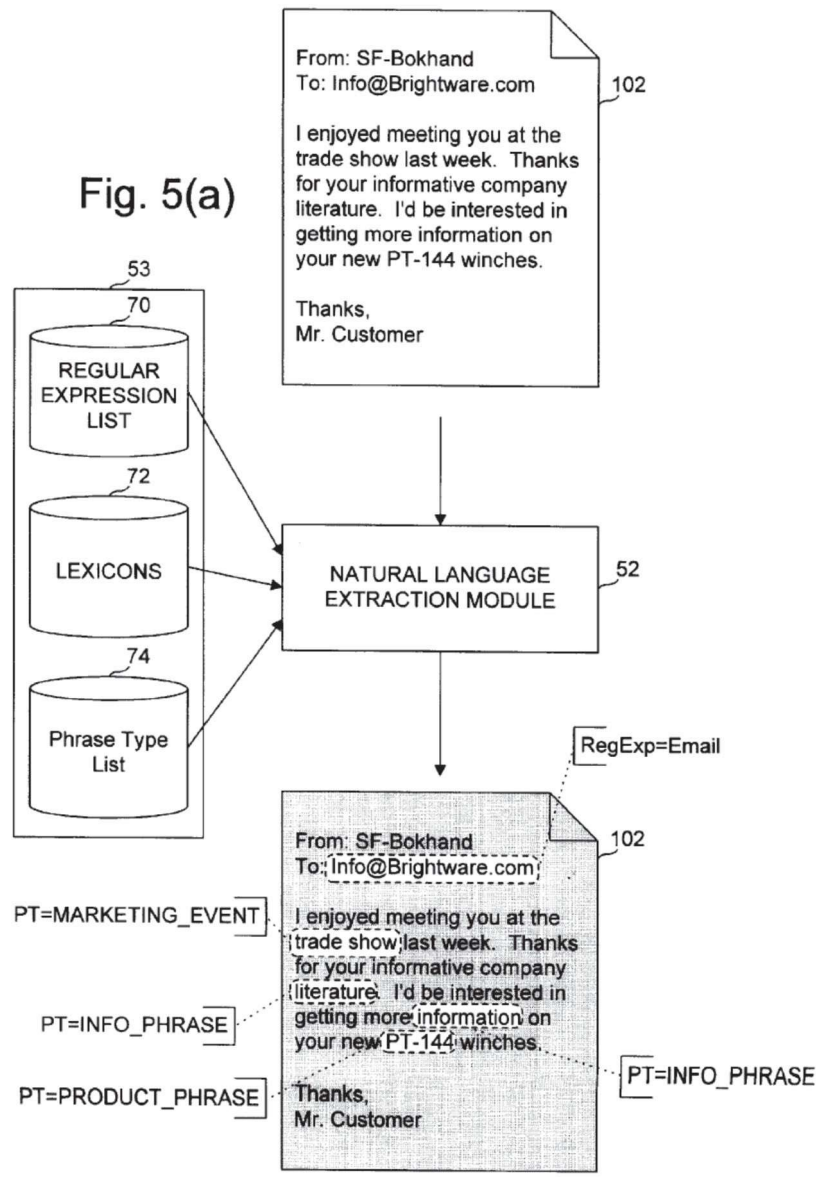
Id., 8:59-66. Richardson explains as follows:

Regular expression list 104 is used to isolate and identify character strings in the message which match the expressions in the list.... In the example shown, an e-mail address is matched and tagged with a “RegExp=Email” tag. Of course, in an actual application, more complex regular expressions could be listed, such as a regular expression that identifies not only that it is an e-mail address, but that it is the “To:” address. Other regular expressions will locate telephone numbers, or even the signature lines of an e-mail message.

Id., 9:17-27. Figure 5(a) depicts this process:

Fig. 5(a)





Id., Fig. 5(a).

36. After regular expressions are identified, “the Richardson’s “analysis process continues, as shown in FIG. 5(b),” which involves “database lookups” and other actions based on extracted data. *Id.*, 10:24-30. “[T]agged Richardson explains that “tagged elements of message 102 are passed to action resolution module (‘ARM’) 56” and ultimately to “action execution module (‘AEM’) 57, which

performs the

(‘AEM’) 57, which performs the actions listed on list 122, such as creating a response 130.” *Id.*, 10:64-65, 11:14-

17. “A Richardson further explains that “[a] preferred response 130 includes all or part of the original message,” along with “tags and the values for tagged regular expressions and tagged phrase types.” *Id.*, 11:17-23.

3.C. Madnick (EX1006)

37. I understand that Madnick is prior art to the ’796 patent at least because it issued in 1999 ~~and is prior art under §102(b).~~ EX1006.

38. Madnick states that it “relates to retrieving data from heterogeneous data sources including structured sources and semi-structured sources and, more particularly, extracting data from World Wide Web pages in response to a query phrased in a structured query language.” *Id.*, 1:15-20. ~~Data~~

39. In Madnick, data is retrieved from data sources 104 (e.g., webpages) using data receivers 102. *Id.*, 3:60-4:18. To extract data from webpages, Madnick discloses “wrapper generator 614,” which comprises “a command transmitter 618; and a data retriever 620.” *Id.*, 9:45-49. Madnick further discloses “[s]pecification file 706 for a data source 104 [that] describes the commands that must be transmitted to the data source 104 by the command transmitter 618 in order to interact with the data source 104.” *Id.*, 11:11-12:2. “The data retriever 620 then extracts the data requested by the data receiver 102 from the semi structured data source 104 using the specification file.” *Id.*, 13:24-26.

| 40. An example of Madnick's specification file 706 is in Table 2:

network.Company	\sSIZE=\d>\(<.*?\s+\W\w+\W#
network.Ticker	\s\(((.*?))>\A#
network.Last	Last\\s+(\<.*?\s+\<A#
network.Last__Trade	\(last\strade\s(\<.*?\sEST\W#
network.Low	Day\sRange\\s\<ID>\s*(.*&)\s*\s*\d#
nethworth.High	Day\sRANGE\\s\<TD>\>.*?\s*(.*?S\s+#
network.Change	Change\\s+(\<.*?\)\n#
network.Prev__Close	Prev.\sClose\\s(\<.*?\)\n#
network.Tick-Trend	Tick\sTrend\\s+\W+TD\>(<.*?\</TD>\>#
network.Volume	Volume\\s+(\<.*?\s+\<A#
network.Market	Market\<A>\s+(\<.*?\s+\n#
network.Open	Open:\<A>\s+(\<.*?\s+\<A#
network.Year__Low	52\sweeks\Range\\s+(\<.*?\s-\#
network.Year__High	52\sweeks\Range\\s+\s-\s(\<.*?\n#
network.PE__Ratio	P/E\sRatio\\s+(\<.*?\s+\<A#
network.Latest__Div	Latest\sDiv.\\s+(\<.*?\s*\<A#
network.Annual__Div	Annual\sDiv.\\s+(\<.*?\s*\</PRE>\>#

network.Company	\sSIZE=\d\>(. *?)\s+\W\w+\W#
network.Ticker	\s\(((*?))\>A#
network.Last	Last\</A\>\s+(. *?)\s+\<A#
network.Last__Trade	\(last\strade\s(. *?)\sEST\W#
network.Low	Day\sRange\</A\>\s\<ID\>\s*(. *&)\s*\s*\d#
nethworth.High	Day\sRANGE\</A\>\s\<TD\>. *?)\s*(. *?)S\s+#
network.Change	Change\</A\>\s+(. *?)\n#
network.Prev__Close	Prev.\sClose\<A\>\s(. *?)\n#
network.Tick-Trend	Tick\sTrend\</A\>\s+\W+TD\>(. *?)</TD\>#
network.Volume	Volume\</A\>\s+(. *?)\s+\<A#
network.Market	Market\<A\>\s+(. *?)\s+n#
network.Open	Open\<A\>\s+(. *?)\s+\<A#
network.Year__Low	52\sweeks\Range\</A\>\s+(. *?)\s-#
network.Year__High	52\sweek\sRange\</A\>\s+(. *?)\s+.\s\<A\>\s+n#
network.PE__Ratio	P/E\sRatio\</A\>\s+(. *?)\s+\<A#
network.Latest__Div	Latest\sDiv.\</A\>\s+(. *?)\s*\<A#
network.Annual__Div	Annual\sDiv.\</A\>\s+(. *?)\s*\</PRE\>#

41. Madnick explains that “Table 2 above uses various PERL operators in order to specify the commands which must be issued.” *Id.*, 13:1-2. Madnick describes perl’s syntax and notes that “[t]he meaning of PERL regular expressions are well known in the art.” *Id.*, 13:2-7; *see also id.*, 15:3-10. (illustrating another example of perl’s regular expression syntax). Madnick explains that “[i]t is preferred that the language used have good text-handling capabilities,” with

“PERL” being an expressly preferred language. *Id.*, 16:23-25.

4.D. Groner (EX1007)

I understand that Groner was prior art to the '796 patent at least because it issued from an application filed on March 16, 2000, ~~and is prior art under §102(e).~~

42. . EX1007.

43. Groner states that it “relates “generally to speech recognition systems as applied to voice and electronic message mailing, and particularly to a system

and method for converting speech to a text message suitable for sending as an e-mail message and for viewing on a text display device.” *Id.*, 1:6-10. Groner notes that “[v]oice mail messages ... are convenient because telephones are ubiquitous and inexpensive” but have

disadvantages compared to written messages, which “can be searched easily for content whereas voice messages cannot be as easily searched.” *Id.*, 2:17-20, 2:31-32.

44. ~~32.~~ To solve this problem, Groner discloses a system and method whereby “[a]n audio message from a caller for a recipient is received[,] [a]n e-mail address for the recipient is determined[,] [a] text message file is generated from the audio message from the caller[,] [and] [t]he text message file is sent to the recipient at the recipient’s e-mail address.” *Id.*, 2:63-67; *see also id.*, 4:38-42, Fig. 2.

Groner explains: “Because the voice messages are in text form, the voice messages can be searched for particular content.” *Id.*, 3:37-39.

45. Groner discloses “a computer system 70 [that] implements the voice-to-electronic mail system 30,” which “automatically converts a spoken message to a text message which is e-mailed to a recipient.” *Id.*, 5:5-9. Among other components, the system includes “speech recognition procedure 116 that receives audio speech, identifies the audio speech and generates a text file 118 corresponding to the identified audio speech.” *Id.*, 6:6-8. Groner notes that “[t]he

accuracy of speech recognition software has improved” and provides examples of commercial embodiments. *Id.*, 2:38-49.

IV. Opinions Regarding Obviousness

46. In my opinion, the challenged claims are unpatentable based upon at least the following three grounds. First, claims 1-10, 12-21, 23, 25 would have been obvious over Sheth. Second, claims 1-10, 12-21, 23, and 25 would have been obvious over Richardson and Madnick. Third, claims 11, 22, and 24 would have been obvious over Richardson, Madnick, and Groner.

B.A. Ground 1: Claims 1-10, 12-21, 23, and 25 Would Have Been Obvious Over Sheth (EX1004)

1. Independent Claims 1, 12, 23, and 25

47. In my opinion, independent claims 1, 12, 23, and 25 would have been obvious to a POSITA over Sheth for at least the reasons set forth below.

a. Claim 1

i. “A method of automatically processing an input sequence of data symbols, the method comprising the steps of”

To the extent ~~claim 1’s~~the preamble is a limitation, it is my opinion that Sheth discloses it. ~~EX1003~~

48. ¶73. Sheth describes “a method for creating a database of metadata (metabase) of a variety of digital media content, data sets, including TV and radio content potentially delivered on Internet.” EX1004, 4:54-57. Specifically, Sheth discloses an “extraction process during which metadata for digital media is

discovered, copied from a Web site, enhanced, and inserted into [a] database.” *Id.*, 10:7-10. “CrawlingIn Sheth, “[c]rawling and extraction rules are input to WebCrawler and Extractor programs that traverse Web sites and retrieve digital media metadata from selected pages,” such that the “extractor program takes HTML pages and extraction rules as input and generates XML assets.” *Id.*, 10:40-45.

“WhenSheth discloses that “[w]hen the WebCrawler determines that it has found an extractable page, it sends the contents of that page ... to another remote piece of software (an Extractor) that retrieves the valuable metadata from the page.” *Id.*, 11:7-9. The processed webpages contain extractable portions of text“. See, e.g., id., 11:17-25), which is. A POSITA would have understood that Sheth’s extractable webpages are an embodiment of an input “sequence of symbols” within the meaning of the ’796 patentthat, which includes

49. “a data stream representative of text” ~~EX1001, 1:58-89~~; ~~EX1003~~

~~¶74.~~ Thus, Sheth discloses a method of processing an input sequence of data symbols—i.e., processing text in webpages. ~~Id.~~

50. Moreover, Sheth repeatedly makes clear that its data-processing method is automatic. EX1004, 5:5-7 (describing “a distributed method and apparatus to quickly produce agents which automatically create and manage digital media metadata”), 9:43-45 (extraction scheduled “automatically”), 13:30-31 (“most of the metadata acquisition happens automatically”); ~~EX1003 ¶75.~~

51. Thus, Sheth discloses a method of automatically processing an input sequence of data symbols. ~~EX1003 ¶76.~~

- ii. **“identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression”**

It is also my opinion that Sheth discloses, teaches, or suggests the first clause in the body of claim 1.

Id. ¶77. As Sheth explains: “Once an extractor receives a page, it looks up a set of extraction rules 8 associated with the site from which the page came,” which “list the metadata attributes for the type of media that this site contains as well as rules that describe where to find values for these attributes within the page.” EX1004, 11:41-46. “The extractor scans the Web page content for pieces of text that match the pattern specified by the extraction rules,” ~~and~~

52. .” *Id.*, 11:50-52. Sheth explains that “[t]he ability to retrieve text that matches a particular pattern is provided by the PERL regular expression language.” *Id.*, ~~11:50-52~~, 11:64-66.

53. ~~A~~In my opinion, a POSITA would have understood that, by “scan[ning] ... for pieces of text that match the pattern specified by the extraction rules,” which are defined by perl regular expressions (~~*id.*~~, 11:50-52, 11:64-66), Sheth’s extractors meet claim 1’s step of “identifying at least one regularly identifiable expression

[i.e., the ‘pattern’ specified by the ‘extraction rules’] in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression” (i.e., perl). ~~EX1003 ¶78.~~

iii. “identifying at least a portion of information associated with the at least one regularly identifiable expression; and”

For similar reasons to those discussed above, it is my opinion that Sheth discloses, teaches, or suggests the second clause in the body of claim

54. 1. ~~Id. ¶79.~~ Sheth’s “[e]xtractors 7 are programs that are designed to find information about digital media from a Web page,” ~~and.~~ EX1004, 11:11-12. For example, extraction “rules can be written that reliably retrieve small sections of text from those pages.” ~~EX1004, 11:11-12,~~ Id., 11:18-20. A “typical extraction rule designates the pattern for which the extractor should search to locate the value for [a particular] attribute.” *Id.*, 12:30-

32. Again, “[t]he ability to retrieve text that matches a particular pattern is provided by the PERL regular expression language,” which allow Sheth’s extractors to “pick out relevant information from a Web page.” Id., 11:64-66, 11:60-63 ~~allows Sheth’s.~~

extractors to “pick out relevant information from a Web page.” In my opinion, a~~*Id.*, 11:64-66, 11:60-63; EX1003 ¶79.~~

55. A POSITA would have understood that, by “locat[ing] the value” for an attribute using a matchable pattern (~~EX1004~~*id.*, 12:30-32) and “retriev[ing] small sections of text” (*id.*, 11:18-20), Sheth’s extractors meet claim

1's step of "identifying at least a portion of information associated with the at least one regularly identifiable expression"—i.e., the "value" or "small sections of text" that are identified when a regular expression is matched. ~~EX1003 ¶80.~~

iv. "extracting the portion of information."

56. For similar reasons to those discussed above, it is also my opinion that Sheth discloses, teaches, or suggests the final clause in the body of claim 1. ~~Id. ¶81.~~ Again, Sheth's ~~extraction "rules can be written that reliably retrieve small sections of text from those pages," and the~~ "extractor scans the Web page content for pieces of text that match the pattern specified by the extraction rules." EX1004, 11:~~48-20, 11:~~50-52. "After the extractor has attempted to find every attribute in its list of extraction rules, it creates an XML document containing attribute-value pairs" (i.e., database attributes matched to the extracted portions of information), which "is sent to the Metabase Agent 10." *Id.*, 11:52-55, Fig. 6; *see also id.*, 10:43-49 ("An extractor program takes HTML pages and extraction rules as input and generates XML assets" that "are sent to a Metabase Agent that is in charge of enhancing and inserting them into a database of records."); ~~EX1003 ¶81..").~~

57. In my opinion, a POSITA would have understood that, by “retriev[ing] small sections of text” that match predefined patterns (i.e., regular expressions) and creating “an XML document” with the desired portions of information (*id.*, 11:18-20, 11:52-55), Sheth meets claim 1’s step of “extracting the ~~portion of information~~” i.e., ~~the information associated with a regular expression.~~ EX1003 ¶82.

portion of information”—i.e., the information associated with a regular expression, which is stored in an XML file and database.

58. Thus, it is my opinion that Sheth teaches or suggests all limitations of claim 1 and, accordingly, would have rendered claim 1 obvious to a POSITA. ~~Id.~~

~~¶83.~~

b. Claim 12

59. Claim 12 recites:

12. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:
at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and
memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

60. In my opinion, Sheth renders claim 12 obvious for substantially the same reasons discussed above for claim 1. ~~EX1003 ¶85.~~ The only differences between claims 1 and 12 are that claim 12 recites (i) an apparatus instead of a method; (ii) “at least ~~one processor operative to~~ perform the steps of claim 1; and (iii) ~~a final memory limitation.~~ *Id.* Sheth discloses each of these limitations. *Id.*

one processor operative to” perform the steps of claim 1; and (iii) a final memory limitation. In my opinion, Sheth discloses each of these limitations.

61. First, Sheth is directed to both “*a system* and a method for creating a database of metadata (metabase) of a variety of digital media content,” as well as a “method *and apparatus* to quickly produce agents which automatically create and manage digital media metadata.” EX1004, 4:54-57, 5:5-7 (emphasis added); *see also id.*, 8:11-12 (~~refereneing~~referring to “the apparatus by which metadata is acquired”), 15:51-52 (“The methods and apparatus of the present invention....”), 17:55-65, 19:60-20:12 (reciting invention as system). Thus, Sheth renders obvious the invention discussed in claim 1 in the form of an apparatus. ~~EX1003 ¶86.~~

62. Second, Sheth discloses a processor operative to perform the claimed functions. ~~Id. ¶87.~~ Sheth explains that “the present invention is directed to a system for generating metadata associated with a data set, the system comprising: *a processor*,” which among other functions “is programmed to ... extract metadata associated with the data set based upon the retrieved extraction information; and storing in the data store the extracted metadata.” EX1004, 17:55-65 (emphasis added). Thus, in my opinion, Sheth discloses a processor that is operative to perform the recited functions.

~~added). Thus, Sheth discloses a processor operative to perform the recited functions. EX1003 ¶87.~~

63. Third, Sheth discloses claim 12's memory limitation. ~~Id. ¶88.~~ In addition to a processor, Sheth's system comprises "a data store coupled to the processor," which

is used to “store the extracted metadata.” EX1004, 17:57-64. The ’796 patent admits that “memory” broadly encompasses various data stores, including “any memory devices associated with a processor or CPU, such as, for example, RAM, ROM, a fixed memory device (e.g., hard drive), a removable memory device (e.g., diskette), flash memory, etc.” EX1001, 12:35-39. ThusAccordingly, in my opinion, Sheth discloses memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations. ~~EX1003 ¶88.~~

64. AccordinglyThus, in my opinion, claim 12 would have been obvious. ~~Id.~~
¶89 to a POSITA over Sheth.

c. Claim 23

65. Claim 23 recites:

23. An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression;

~~accordance with a programming language that supports such a
regularly identifiable expression;~~

identifying at least a portion of information associated with the
at least one regularly identifiable expression; and

extracting the portion of information.

66. In my opinion, Sheth renders claim 23 obvious for substantially the same reasons as claims 1 and 12. ~~EX1003 ¶91~~. The only difference between claims 1 and 23 is the preamble: Instead of a method, claim 23 is directed to an article of manufacture comprising a machine-readable medium containing one or more programs which when executed implement the same steps as claim 1. ~~Id.~~ To the extent this preamble is limiting, Sheth discloses it. ~~Id.~~

67. Sheth states that it “is directed to *software*, a system and a method for creating a database of metadata (metabase)” for Internet media content. EX1004, 4:54-57 (emphasis added); *see also id.*, 6:6-7 (discussing “a software architecture according to the present invention”), 10:57-58, 11:8-9 (describing both WebCrawler 3 and Extractors 7 as “a piece of software”), 17:21-22 (referring to them as “software components”). Moreover, as discussed for claim 12, Sheth discloses a “system” comprising “a processor, and a data store coupled to the processor and storing WorldModel-based extraction information,” wherein “[t]he processor is *programmed* ... to extract metadata” and for “storing ... the extracted metadata,” among other claimed functions. *Id.*, 17:55-65 (emphasis added). ~~... the extracted metadata.” *Id.*, 17:55-65 (emphasis added).~~

68. Thus, in my opinion, Sheth discloses an article of manufacture comprising a machine-readable medium containing one or more programs which when executed implement the claimed steps, rendering claim 23 obvious.

EX1003 ¶93.

d. Claim 25

69. Claim 25 recites:

25. A method of automatically processing an input document, the method comprising the steps of:

identifying one or more regular expressions in the input document by comparing the input document to one or more previously--stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions, wherein a regular expression represents a pattern that is matchable in accordance with a programming language that supports such a regular expression; and
identifying at least a portion of information associated with the at least one regular expression for extraction.

In my opinion, Sheth renders claim 25 obvious for substantially the same reasons as claim

70. 1. ~~EX1003 ¶95.~~ The only difference between claims 1 and 25 is that claim 25 omits the final step in claim 1 that requires extracting information and, instead, specifies that regular expressions are identified by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions. Sheth discloses this additional element.

~~match between a portion of the input document and at least one of the previously-stored regular expressions. *Id.* Sheth discloses this additional element. *Id.*~~

71. Sheth explains: “~~Once that~~ “[o]nce an extractor receives a page, it looks up a *set of extraction rules* 8 associated with the site from which the page came” and “scans the Web page content for *pieces of text that match the pattern specified by the extraction rules.*” EX1004, 11:41-52 (emphasis added). “The ability to retrieve text that *matches a particular pattern* is provided by the PERL regular expression language.” *Id.*, 11:64-66 (emphasis added).

72. ~~A~~In my opinion, a POSITA would have understood that Sheth’s extraction rules comprising regular expressions are previously stored. ~~EX1003 ¶97.~~ ~~Before~~Sheth discloses that, before the automatic extraction process begins, “a human writes a set of extraction rules for a site.” EX1004, 11:35-36. The extractor then “looks up [the] set of extraction rules” for a particular ~~site~~website and uses the retrieved “list of extraction rules” to find desired attributes. *Id.*, 11:41-43, 11:52-53. Similarly, Sheth explains that “extraction rules are written” first and then these “extraction rules are input to ... Extractor programs.” *Id.*, 10:40-41; *see also id.*, 10:43-44 (“An extractor program takes ... extraction rules as input”). Accordingly, a POSITA would have understood that Sheth’s extraction rules comprising regular expressions are stored (in a “list of extraction rules” that the extractors “look[] up”) prior to identifying the regular expressions in the input document webpage. ~~EX1003~~

~~¶97.~~

73. Thus, in my opinion, claim 25 would have been obvious. ~~Id.~~ ¶ 98 to a
POSITA over Sheth.

2. Dependent Claims 2-10 and 13-21

74. It is also my opinion that the following dependent claims of the '796 patent would have been obvious to a POSITA over Sheth.

a. Claims 2 and 13

75. Claims 2 and 13 depend from claims 1 and 12, respectively, and addfurther recite that identifying the regularly identifiable expression “*comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.*”

This limitation is substantively identical to the limitation added by claim 25 and, for the same reasons discussed abovein the preceding section, is disclosed by Sheth. *Supra* §V.BVII.A.1.d; EX1003

76. ¶101. Thus, in my opinion, claims 2 and 13 would have been obvious to a POSITA over Sheth. EX1003 ¶101.

b. Claims 3 and 14

77. Claims 3 and 14 depend from claims 1 and 12, respectively, and further recite “*normaliz[ing] the input sequence of data symbols prior to identifying the regularly identifiable expression.*”

78. In my opinion, Sheth discloses, teaches, or suggests this limitation. EX1003 ¶103. for

at least two reasons. First, Sheth’s “process starts by identifying a source containing digital media data sets” and “determin[ing] whether a source is actually

worth extracting.” EX1004, 10:23-25. To do so, Sheth uses “WebCrawler 3,” whose purpose “is to read in the content of ‘extractable’ Web pages from a site and pass this content on to the Extractors.”

Id., 10:55-57. WebCrawler 3 “begins reading pages from a particular site and determines which of these pages are extractable.” *Id.*, 10:57-60. WebCrawler 3 invokes “crawling rules” that “define the characteristics of an extractable page.” *Id.*, 10:60-65. “When the WebCrawler determines that it has found an extractable page, it sends the contents of that page 6 on to another remote piece of software (an Extractor) that retrieves the valuable metadata from the page.” *Id.*, 11:7-10.

79. Sheth’s process of using a WebCrawler to sort between extractable and unextractable webpages and passing only the extractable webpages to the Extractors is a process of normalizing data. ~~EX1003-¶104.~~ The ’796 patent explains that one example of normalization “may include standardizing any characteristics of the input that are irrelevant to, or a hindrance to, information extraction.” EX1001, 4:12-15. Similarly, Sheth standardizes input data from webpages by excluding pages that are unextractable and thus a hindrance to information extraction. ~~EX1003-¶104.~~ Because this normalization takes place before webpages are sent to the Extractors, a POSITA would have understood it occurs prior to identifying the regularly identifiable expression, satisfying claims 3 and 14. ~~*Id.*~~

~~¶104.~~

~~Moreover~~

Second, Sheth discloses an additional type of normalization by “convert[ing] the format of an extracted value into a standard format”—e.g., “convert[ing] a clip length in minutes:seconds to seconds (the format accepted by the metabase).” EX1004, 12:50-54; ~~EX1003 ¶105.~~ Although this normalization

80. occurs after identifying regular expressions in Sheth, it would have been equally obvious to perform it ~~beforehand. EX1003 ¶105.~~ Before identifying any regular expressions. Specifically, a POSITA would have recognized that normalizing data to a standard format before identifying regular expressions would simplify the process of writing extraction rules, which Sheth teaches should be designed to “consistently retrieve the correct text” across “several different pages.” ~~EX1004, Id.,~~ 12:18-22. Because webpages might recite the same attribute in different formats (e.g., defining a runtime in either minutes:seconds or seconds), normalizing the webpage data to a uniform format would have made it easier for a POSITA to write a single regular expression that consistently retrieves the correct information regardless of format. ~~EX1003 ¶105.~~

81. At a minimum, in my opinion, practicing Sheth’s step of standardizing data formats before identifying regular expressions would have been obvious to try ~~because. I have been informed that an invention would have been obvious to try~~ when there was a design need or market pressure to solve a problem and there were a finite number of identified, predictable solutions. Here, the problem to be solved is standardizing data for entry in a database, and there were only two ~~possibilities~~ possible

solutions for when to practice Sheth’s normalization step—i.e., normalizing webpage data before or after identifying regular expressions. *Id.* ¶106; see also *Uber Techs., Inc. v. X One, Inc.*, 957 F.3d 1334, 1340 (Fed. Cir. 2020) (whether to perform a step “before” or “after” another is “A POSITA implementing Sheth thus would have faced a simple design choice” that would have been obvious to try). as to whether to standardize data before or after identifying regular expressions, rendering either option obvious to try.

82. Thus, in my opinion, claims 3 and 14 would have been obvious. EX1003 ¶107 to a POSITA over Sheth.

c. Claims 4 and 15

83. Claims 4 and 15 depend from claims 1 and 12, respectively, and ~~add~~further recite “*identify[ing] one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.*”

84. In my opinion, Sheth discloses this limitation for substantially the same reasons discussed above for claims 3 and 14. ~~EX1003 ¶109; supra §V.B.~~Supra §VII.A.2.b. AgainAs discussed, before Sheth’s Extractors identify any regular expressions, WebCrawler 3 “begins reading pages from a particular site and determines which of these pages are extractable.” EX1004, 10:57-60. Only pages deemed extractable by the WebCrawler are passed to the Extractors. *Id.*, 11:7-10. Sheth explains: ~~“An~~ that “[a]n ‘extractable page’ ... follows a template (however loosely) and likely contains a link to a media asset.” *Id.*, 11:1-3. ~~“A~~Moreover, “[a] crawler can often recognize an extractable page by simply examining the URL. If the URL of the page being

examined follows a certain pattern, there is a good chance that the page contains a link to an audio or video asset.” *Id.*, 11:3-7.

85. Accordingly, Sheth’s WebCrawler identifies one or more classes of data symbols in the input sequence (e.g., text in the URL and/or other characteristic text of extractable pages), which occurs “prior to the Extractors identifying ~~regular~~the regularly identifiable expressions.” Thus, in my opinion, claims 4 and 15 would have been obvious to a POSITA over Sheth.
EX1003 ¶110. Thus, claims 4 and 15 would have been obvious. *Id.*

d. Claims 5 and 16

86. Claims 5 and 16 depend from claims 1 and 12, respectively, and additionally recite: “wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.”

87. In my opinion, Sheth discloses, teaches, or suggests this limitation.
EX1003 ¶112.

For example, Sheth provides examples of extraction rules in Figure 8 below, which use the perl regular expression language. EX1004, 11:60-66.

```
# Name : CNNUS
#
producer          [c]CNN
description        [i]{?i}<img src[^>]*alt="Video"[^>]*>([<]*)<
descriptionReplace1 s/\([^\(\)]*\)/ /
parentURL         [u](.*)
invalidated       [c]0
surrogate         [i]{?i}([<]*)</h>
productionSource  [c]tv
location          [S]{?i}/date.*?<p>([<]*)\([<]*)\
locationReplace1  s/.?<I>.*?<\/I>([<]*)/$1/
locationReplace2  s/([<]*)--.*?/$1/
locationReplace3  s/.?<em>.*?<\/em>([<]*)/$1/
mediaType        [c]video
source           [c]http://www.cnn.com
```

```

# Name : CNNUS
#
producer          [c]CNN
description        [ ](?i)<img src[^>]*alt="Video"[^>]*>{[^<]*}<
descriptionReplace1 s/\\([^\s])*\s\\//
parentURL         [u](.*)
invalidated       [c]0
surrogate         [ ](?i){[^<]*}</h.>
productionSource  [c]tv
location          [S](?i)/date.*?<p>(.*)\\(.*?\\)
locationReplace1  s/. *?<I>.*?<\\I>(.*)/$1/
locationReplace2  s/(.*)--.*/$1/
locationReplace3  s/. *?<em>.*?<\\em>(.*)/$1/
mediaType         [c]video
source            [c]http://www.cnn.com

```

Id., Fig. 8; [EX1003 ¶112](#).

A POSITA would have understood that Sheth’s Figure 8 includes multiple regular expressions wherewherein a characteristic phrase typically precedes a portion of information. [EX1003 ¶113](#). For example, the attribute “surrogate” shown above includes the string “<img src=”, wherein the HTML keywords “img” and “src” form a phrase that typically precedes a portion of an extractable URL for

88. an image. ~~Id.~~ This reflects commonly understood syntax for regular expressions, with which a POSITA would have been familiar. ~~Id.~~

89. Thus, in my opinion, claims 5 and 16 would have been obvious. ~~Id.~~
¶114 to a POSITA over Sheth.

e. Claims 6 and 17

90. Claims 6 and 17 depend from claims 1 and 12, respectively, and further ~~addrecite~~: “*wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.*”

91. In my opinion, Sheth discloses, teaches, or suggests this limitation. ~~EX1003 ¶116.~~ In Figure 8 of Sheth (reproduced above, in ¶112), the regular expressions for each of the “urlReplace[1-5]” attributes includes a phrase that typically follows a particular portion of information. ~~Id.~~ For example, in “urlReplace5,” the regular expression includes the string “/.?*qtref”, where “qtref” follows a portion of ~~a URL. Id.; EX1004, Fig. 8~~ the URL to be retrieved. Id.; EX1004, Fig. 8. Again, this reflects commonly understood syntax for regular expressions, with which a POSITA would have been familiar.

92. Thus, in my opinion, claims 6 and 17 would have been obvious. ~~EX1003 ¶117~~ to a POSITA over Sheth.

f. Claims 7 and 18

93. Claims 7 and 18 depend from claims 1 and 12, respectively, and further ~~addrecite~~: “*wherein the extracted portion of information is used to take a specified action.*”

94. In my opinion, Sheth discloses this limitation. ~~EX1003 ¶119.~~

Sheth's "extractor program takes HTML pages and extraction rules as input and generates XML assets," which "are sent to a Metabase Agent that is in charge of enhancing and ~~inserting them into a database of records.~~" EX1004, 10:43-49; ~~see also id., 11:52-55.~~ Sheth's use

inserting them into a database of records.” EX1004, 10:43-49; see also id., 11:52-

55. Sheth’s use of extracted information to generate XML pages, enhance data assets, and their insertion into a database are all examples wherein the extracted portion of information is used to take specified actions. ~~EX1003 ¶119.~~

95. Thus, in my opinion, claims 7 and 18 would have been obvious. ~~Id.~~
¶120 to a POSITA over Sheth.

g. Claims 8 and 19

96. Claims 8 and 19 depend from claims 1 and 12, respectively, and further recite: “*wherein the extracted portion of information is at least one of visually and audibly presented to the user.*”

97. In my opinion, Sheth discloses this limitation. ~~EX1003 ¶122.~~ Sheth is expressly directed to “extracting metadata associated with [a] data set ... and accessing content information based on the metadata *for presentation to a user.*” EX1004, 17:47-54 (emphasis added); *see also id.*, 19:28 (claiming “presentation to a user”).

98. Sheth For example, Sheth states that its invention “allow[s] users to search the metabase and make use of the semantic information gained in performing” the extraction steps discussed above, such that users obtain “meaningful results.” *Id.*, 13:47-50, 13:56-60. Sheth also discloses interface “skins” that present information to users by topic. *Id.*, 15:12-30. For example,

after collecting URLs using the extraction process discussed in prior steps (*supra*
§~~V.B~~VII.A.1.a), ~~Sheth~~Sheth's invention

may “*present the user* with links to local Internet radio stations grouped by genre.”

EX1004, 15:18-20 (emphasis added). [Sheth further discloses that this information](#)

[is visually presented to the user, as shown by the visual interfaces depicted in](#)

[Figures 13A-C. *Id.*, 15:12-30, Figs. ~~This~~ 13A-C.](#)

information is visually presented to the user, as shown by the visual interfaces in Figures 13A-C. ~~Id., 15:12-30, Figs. 13A-C; EX1003 ¶123.~~

99. Thus, in my opinion, claims 8 and 19 would have been obvious. ~~EX1003 ¶124~~ to a POSITA over Sheth.

h. Claims 9 and 20

100. Claims 9 and 20 depend from claims 1 and 12, respectively, and ~~add~~further recite: “*wherein the regularly identifiable expression identifying [step/operation] is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.*”

101. In my opinion, Sheth discloses this limitation. ~~EX1003 ¶126.~~ Sheth teaches: “The that “[t]he ability to retrieve text that matches a particular pattern is provided by the PERL regular expression language.” EX1004, 11:64-66; *see also id.*, 6:45-47.

102. Thus, in my opinion, claims 9 and 20 would have been obvious. ~~EX1003 ¶127~~ to a POSITA over Sheth.

i. Claims 10 and 21

103. Claims 10 and 21 depend from claims 1 and 12, respectively, and ~~add~~further recite: “*wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence*

data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.”

In my opinion, Sheth discloses this limitation. ~~EX1003 ¶129~~. As discussed above, Sheth processes and extracts data from webpages ~~comprising~~that comprise text. EX1004, 11:50-52

104. (“The extractor scans the Web page content for pieces of text that match the pattern specified by the extraction rules.”); *see also id.*, 11:64-66. Accordingly, in my opinion, the input sequence of data symbols in Sheth is representative of text data, rendering claims 10 and 21 obvious to a POSITA over Sheth. ~~EX1003 ¶129.~~

C.B. Ground 2: Claims 1-10, 12-21, 23, and 25 Would Have Been Obvious Over Richardson (EX1005) and Madnick (EX1006)

105. As an independent basis for my conclusion that claims 1-10, 12-21, 23, and 25 of the '796 patent are unpatentable as obvious, it is also my opinion that these claims would have been obvious to a POSITA in view of Richardson and Madnick. Below, I first explain my opinion on why a POSITA would have combined Richardson and Madnick. I then explain why the Richardson-Madnick combination meets all limitations of claims 1-10, 12-21, 23, and 25.

1. Richardson-Madnick Combination

A POSITA would have recognized that Richardson and Madnick are similar in that both teach the use of regular expressions for identifying and extracting information from documents. ~~EX1003 ¶131; *supra*~~

106. §§V.A.2-3; *Supra* §§VI.B-C.; EX1005, 9:18-20 (“Regular expression list 104 is used to isolate and identify character strings in the

message which match the expressions in the list.”), Figs. 3-4; EX1006 13:24-26 (“[D]ata retriever 620 then extracts the data requested by the data receiver 102 from the semi structured data source 104 using the specification file.”), col. 12, Tbl. 2 (listing regular expressions).

107. A POSITA also would have recognized that Richardson’s process for identifying and extracting information from documents using regular expressions usesis carried out using a programming language running on server 16, which is “programmed according to [its] use” and stores “programs used to implement the programming.” EX1005, 5:11-13, 5:17-19; *see also id.*, 5:4-6 (server 16 comprises knowledge base 36, message processing module 38), 5:23-28 (knowledge base 36 comprises extraction rule set 50; message processing module 38 comprises natural language extraction (NLE) module 52), 5:29-33 (NLE

module 52 comprises regular expression module 60; extraction rule set 50 comprises regular expression lexicon 70), Figs. 1-5; ~~EX1003 ¶132~~.

108. Although Richardson teaches a programmed system and method using regular expressions, Richardson does not explicitly identify which programming language is used. ~~EX1003 ¶133. This to define such regular expressions. In my opinion, this~~ would have led a POSITA to search for and locate Madnick, which expressly discloses that regular expressions are written in perl. ~~Id.~~ Specifically, Madnick discloses that the exemplary regular expressions in Table 2 are “PERL operators” (EX1006,

13:1-2) and recognizes that “PERL regular expressions are well known in the art” (*id.*, 13:2-7). Moreover, Madnick encourages using perl, which is highlighted as a “preferred” language ~~because of~~due to its “good text-handling capabilities.” *Id.*, 16:23-25; ~~EX1003 ¶133~~.

As ~~evidenced~~shown by Madnick, a POSITA would have understood that Richardson’s regular expressions represent patterns that are matchable in accordance with a programming language that supports such a regularly identifiable expression, e.g., perl. ~~EX1003 ¶134~~. Further, in my opinion, Madnick would have motivated a POSITA, with a reasonable expectation of success, to use perl for Richardson’s regular expressions by (i) teaching its use for the same purpose of identifying information in documents (EX1006, 13:1-7); (ii) providing examples of perl regular expressions and their syntax (*id.*, col. 12, Tbl. 2; *id.*, 13:64-67, 15:1:10); and (iii) teaching a preference for perl ~~due to~~because of its text-handling abilities

109. (*id.*, 16:20-25). ~~EX1003-¶134.~~ Using perl as taught by Madnick requires no modification of Richardson and is consistent with Richardson’s objective of extracting textual information using regular expressions. ~~Id.~~ Moreover, the ’796 patent admits that perl has “built-in support for regular expressions” and has been ~~well-~~known in the art since at least 1996. EX1001, 2:26-27, 3:23-31. Thus, a POSITA would have been motivated to combine Richardson and Madnick for this limited purpose with a reasonable expectation of success. ~~EX1003-¶134.~~

2. Independent Claims 1, 12, 23, and 25

110. As shown below, it is my opinion that Richardson ~~as evidenced by, or in~~
~~view of, and~~ Madnick ~~meets~~meet all elements of ~~the following~~independent claims 1, 12,
23, and 25 of the '796 patent.

a. Claim 1

i. “A method of automatically processing an input sequence of data symbols, the method comprising the steps of”

To the extent the preamble of claim 1 is ~~limiting~~a limitation, Richardson discloses it. ~~EX1003~~

111. ~~¶136.~~ The '796 patent acknowledges that “symbols” include “letters” and “characters,” including in “a data stream representative of text.” EX1001, 1:58-59. ~~Richardson discloses a method of processing such text symbols. EX1003 ¶137.~~

In my opinion, Richardson discloses a method of processing such text symbols. Specifically, Richardson “is directed to a method and apparatus for automatically recognizing and responding to intent in message text.” EX1005, 1:21-26. “The system recognizes phrase types and intents from input text and maps these to domain knowledge pertaining to common kinds of intents.” *Id.*, 4:15-18. “The

112. input message arrives at the system via electronic mail, network transport, or any number of well-known communication methods.” *Id.*, 4:23-25. “FIGS. 5(a)-(b) show an example of an input message 102 being processed....” *Id.*, 8:42-43. “The input message is processed by NLE module 52 to generate tags associated with portions of the input message.” *Id.*, 8:43-44.

113. Thus, in my opinion, Richardson discloses a method of automatically processing an input sequence of data symbols. ~~EX1003 ¶138.~~

- ii. **“identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression”**

114. In my opinion, Richardson, as evidenced by or in view of and Madnick, teaches teach the first clause in the body of claim 1. EX1003 ¶139 of the '796 patent.

First, it is my opinion that Richardson’s process includes identifying a regularly identifiable expression. Id. ¶140. “TheIn Richardson, “[t]he process of understanding and responding to text messages begins with accepting the input text message, which is first processed by NLE module 52.” EX1005, 5:34-37. Within Richardson’s NLE module 52, “regular expression module 60 receives the input text message and reads from regular expression lexicon 70.” Id., 5:37-40, Fig. 4. “RegularRichardson explains that “[r]egular expression module 60 looks for regular expressions which are known to regular expression lexicon 70.” Id., 5:51-53.

115. “Once a regular expression[] is identified, it is excluded from further analysis in the free-form text.” *Id.*, 5:60-61 (emphasis added). Thus, Richardson discloses identifying at least one “regularly identifiable expression.” ~~EX1003 ¶140.~~”

Second, in my opinion, Richardson’s regular expressions represent patterns that are matchable ~~patterns.~~ *Id.*

116. ~~¶141.~~ For example, “[r]egular expression list 104 [that] is used to isolate and identify character strings in the message *which match the*

TABLE 1

Regular Expression List	
Expression	Type
http://*{/*}>	URL
###-###-####	Phone
[a-zA-Z]@*.*	Email
. . .	. . .

expressions in the list.” EX1005, 9:17-19 (emphasis added). Moreover, [Richardson explains that](#) “after the spelling module removes a typographical error, the result may *match a regular expression* which was not earlier matched....” *Id.*, 6:14-19 (emphasis added). Likewise, Richardson’s claim 13 recites “*matching* identified tokens to a database of *regular expression patterns*.” *Id.*, 12:53-54 (emphasis added). A POSITA would also have understood that Richardson’s regular expressions represent matchable patterns, such as the examples in Table 1 [below](#) that match patterns for a URL, phone number, or email address. ~~EX1003~~

TABLE 1

Regular Expression List	
Expression	Type
http://*{ /* }>	URL
###-###-####	Phone
[a-zA-Z]@*, *	Email
...	...

[¶141.](#)

EX1005, 8:59-66. Thus, Richardson teaches identifying “at least one regularly identifiable expression that represents a pattern that is matchable. ~~EX1003~~ [¶141.](#)”

Third, in my opinion, a POSITA would have understood that Richardson's regular expressions represent a pattern matchable in accordance with a programming language that supports such a regularly identifiable expression. ~~Id.~~ ¶142.

117. Richardson's method of extracting information from input text by identifying regular expressions is implemented using a programming language.

~~Id.~~ In particular, regular expression matching in Richardson is performed by server

16, which is a

~~Specifically, regular expression matching is performed by server 16—a~~

“general purpose computer system[] programmed according to [its] use,” for which “[t]he programs used to implement the programming might reside in memory” or other storage devices. EX1005, 5:11-19; ~~EX1003 ¶142.~~ Richardson’s NLE module 52, which comprises regular expression module 60 (EX1005, 5:29-30), is part of processing module 38 (*id.*, 5:23-28), which in turn is part of server 16 (*id.*, 5:4-6). *See also* Fig. 1 (server 16), Fig. 2 (server 16 comprises message processing module 38), ~~Fig. (server 16), Fig. 2 (server 16 comprises message processing module 38), Fig. 3 (message processing module 38 comprises NLE module 52), Fig. 4 (NLE module 52 comprises regular expression module 60); EX1003 ¶142.~~

118. Thus, a POSITA would have understood that regular expression module 60, which “looks for regular expressions” and “identifies instances of the regular expressions” (EX1005, 5:51-59), functions in accordance with the programming of server 16, where regular expression module 60 resides ~~(EX1003 ¶143).~~ A. Accordingly, in my opinion, a POSITA would have understood that server 16 is necessarily programmed with a “a programming language that supports such a regularly identifiable expression” because server 16 comprises regular expression module 60.

~~programming language that supports such a regularly identifiable expression
because server 16 comprises regular expression module 60. EX1003 ¶143.~~

119. ~~A~~Further, a POSITA would also have understood that Richardson's regular expressions represent patterns matchable in accordance with "a programming language that supports such a regularly identifiable expression" based on Madnick. ~~Id. ¶144.~~ Madnick acknowledges that "[t]he meaning of PERL regular

expressions are well known in the art” and recites “various PERL operators” to match regular expressions. EX1006, 13:6-7, 13:1-2, col. 12, Tbl. 2; *see also id.*, 15:1-10 (discussing perl syntax for regular expressions); *Realtime Data, LLC v. Iancu*, 912 F.3d 1368, 1373-74 (Fed. Cir. 2019) (holding “the Board had no obligation to find a motivation to combine” where petitioner “relied on [a second reference] simply to demonstrate that a person of ordinary skill in the art would have understood that [an embodiment] disclosed in [a first reference] was, in fact, a type of [limitation using] the terminology used in the [challenged] patent”). Thus, to a POSITA reading the reference, Richardson would at least implicitly disclose “a programming language that supports a regularly identifiable expression.”

120. At a minimum, a POSITA practicing Richardson would at least have found it obvious to use a programming language that supports a regularly identifiable expression, such as perl. ~~EX1003 ¶145.~~ in view of Madnick. As I discussed above in ~~§V.A.3 and §V~~ §VI.C.1, Madnick encourages ~~using a POSITA to use~~ perl for coding regular expressions as a “preferred” language with “good text-handling capabilities.” EX1006, 16:22-24; ~~EX1003~~ ¶145.24. Again, Madnick repeatedly illustrates the use of perl for coding regular expressions

in the same context of extracting textual information as Richardson. EX1006, col. 12, Tbl. 2, 13:1-7, 15:1-10; ~~EX1003 ¶145.~~ Thus, in my opinion, a POSITA would have ~~combined been motivated to combine~~ Madnick's use of perl with Richardson's process ~~for matching to match~~ a regular expression in accordance with "a programming language that supports such a regularly identifiable expression. ~~EX1003 ¶145; see also Realtime Data, 912 F.3d at 1373-74 (holding in the alternative that a POSITA would have been motivated to combine a secondary reference that was cited as evidencing a teaching in a first reference).~~"

121. Moreover, I note that the '796 patent admits that the claimed regularly identifiable expressions may be "stored in one of several common programming languages, e.g., as a 'flex' or 'perl' program," which "have built-in-support for regular expressions." EX1001, 2:23-27. The '796 further admits that regular

expressions include “any form of a pattern that is matchable in the flex, lex, or perl programming languages,” which were admittedly known in the art. *Id.*, 3:23-31.

~~Accordingly, the '796 patent concedes that regularly identifiable expressions are, by definition, matchable in accordance with a programming language that supports such a regularly identifiable expression. EX1003 ¶146.~~

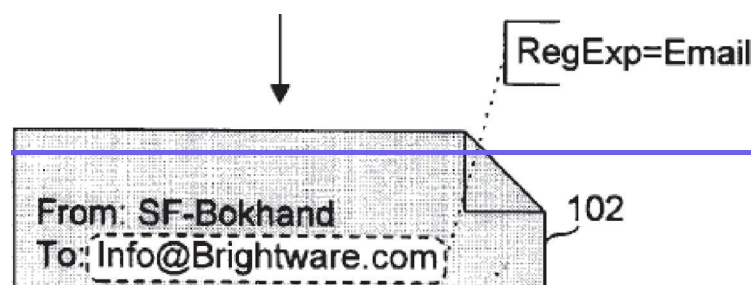
Accordingly, the '796 patent itself concedes that regularly identifiable expressions are, by definition, matchable in accordance with “a programming language that supports such a regularly identifiable expression.”

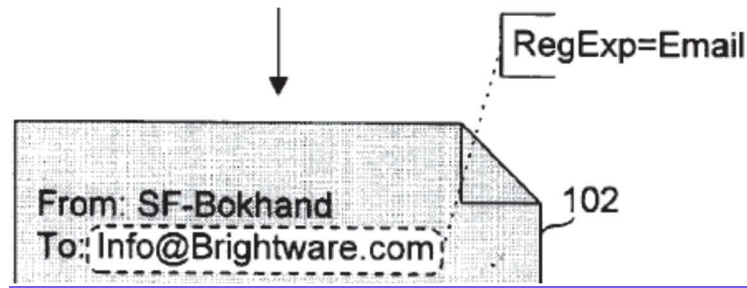
122. Thus, in my opinion, Richardson ~~as evidenced by, or in view of, and~~ Madnick ~~teaches~~teach the first clause in the body of claim 1. ~~*Id.* ¶147 of the '796 patent.~~

iii. “identifying at least a portion of information associated with the at least one regularly identifiable expression; and”

123. It is also my opinion that Richardson further discloses the second clause in claim 1’s body. Id. ¶148. “As Richardson disclose that, “[a]s shown in FIG. 5(a), NLE module 52 uses a regular expression list 104 ... in assigning tags to a message.” EX1005, 8:50-52. “Regular Richardson explains that “[r]egular expression list 104 is used to *isolate and identify character strings* in the message *which match the expressions* in the list.” *Id.*, 9:17-19 (emphasis added). “If a regular expression from list 104 is matched in the message, the matching text is tagged with a ‘RegExp’ tag.” *Id.*, 9:19-21.

124. In Table 1, an exemplary regular expression to identify email addresses is “[a-zA-Z]@*.*]”. *Id.*, 1:64. In Figure 5(a) excerpted below, “an e-mail address is matched and tagged with a ‘RegExp=Email’ tag.” *Id.*, 9:21-22.





Id., Fig. 5(a) (detail). In this example, the “regularly identifiable expression” within the meaning of claim 1 of the ’796 patent is “[a-zA-Z]@*.*” and the portion of information associated with the regular expression is the specific email address “Info@Brightware.com”. ~~EX1003 ¶149.~~

125. Thus, in my opinion, Richardson discloses identifying “at least a portion of information associated with the at least one regularly identifiable expression. ~~*Id.* ¶150,~~” as recited in claim 1 of the ’796 patent.

iv. “extracting the portion of information.”

126. It is also my opinion that Richardson further discloses, teaches, or suggests claim 1’s final clause. ~~*Id.* ¶151.~~ Notably, the module in Richardson that identifies regular expressions is called the “Natural Language *Extraction* Module,” and the regular expressions themselves are labeled “*Extraction* Rules.” EX1005, Figs. 3-5, 3:43-49; *see also id.*, 5:27-30 (“[A] natural language *extraction* (NLE) module 52” is shown in Fig. 4 “comprising a regular expression module 60.”), 5:32-33 (“*Extraction* rule set 50 is shown comprising a regular expression lexicon 70.”) (emphases added). Thus, a POSITA would have understood that Richardson

expressly characterizes regular expressions as a means for extracting matchable portions of information. ~~EX1003 ¶151.~~

Moreover, Richardson discloses specific examples of extracting portions of information identified with regular expressions. ~~Id. ¶152.~~ As I discussed above, one example is an email address that is identified using a regular expression. ~~Id. ¶152.~~ Fig.

127. 5(a), 9:21-22. Richardson discloses that the email address may be extracted for use in a database query. *Id.*, 10:24-30 (“Using the tagged phrases and regular expressions, the system performs any necessary database lookups. For example, a database lookup might be used to determine the customer’s full identity just from a name in the message or the e-mail address.”); ~~EX1003 ¶152.~~ 127.

128. Richardson also discloses that “[t]he message with its tags”—e.g., portions of information matching regular expressions—“are then provided to intent matching module 54, which identifies the writer’s and system operator’s intents using an intent list 114.” EX1005, 10:35-38. Thus, in Richardson, identified information is extracted for transmission to intent matching module 54. ~~EX1003 ¶153.~~

129. In yet another example in Richardson, a reply to the input text is generated that includes extracted information identified by regular expressions:

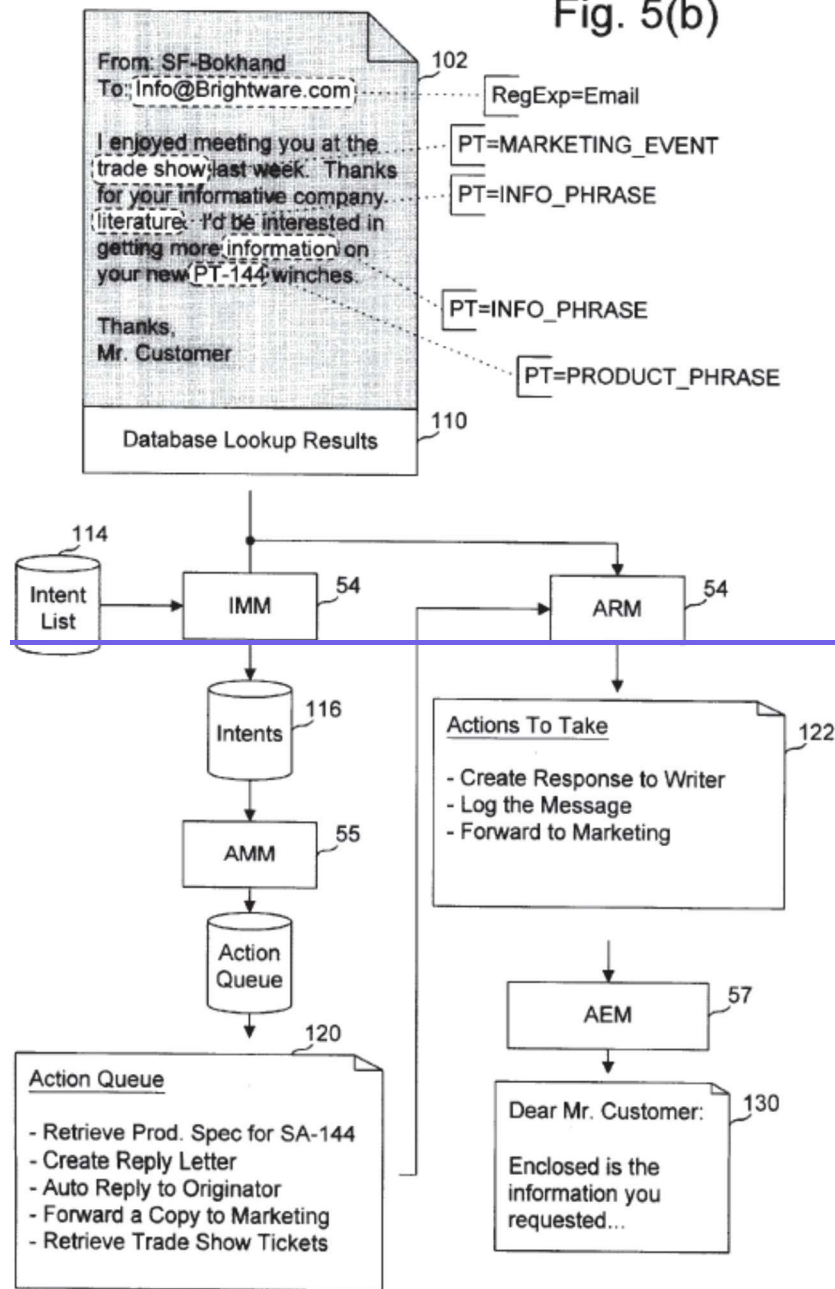
[A] list 122 of actions to take is passed to action execution module (“AEM”) 57, which performs the actions listed on list 122, such as creating a response 130. *A preferred response 130 includes all or part of the original message, ... and an indication of how the message*

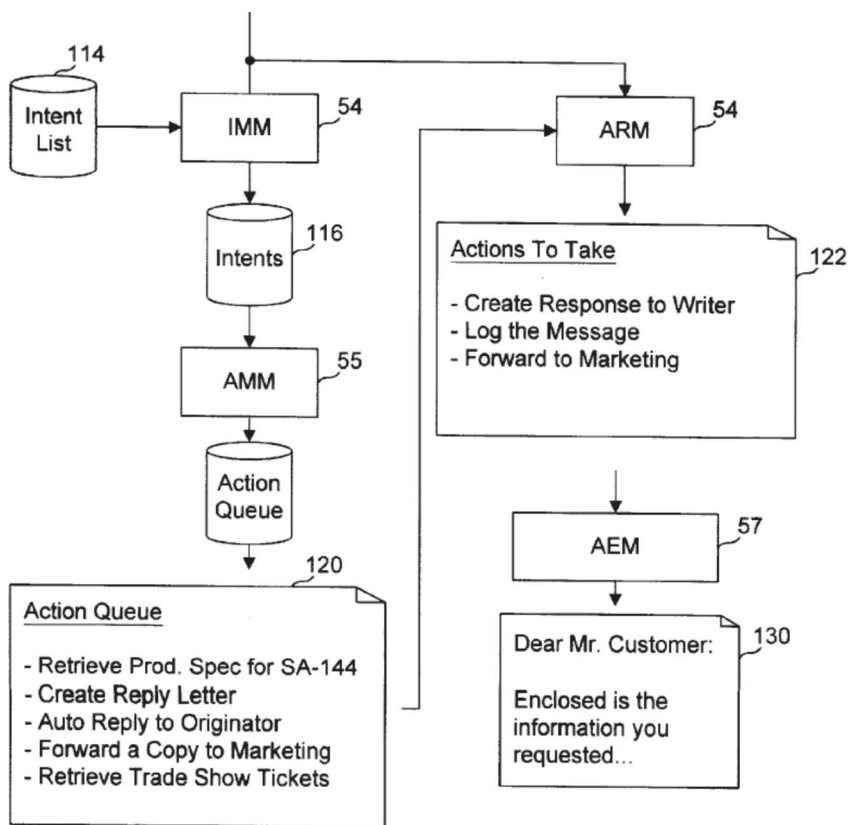
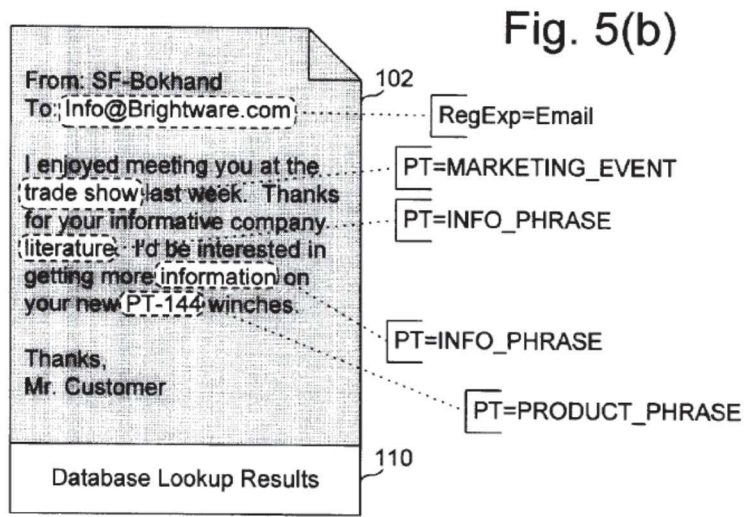
was interpreted. *One way to show how the message was interpreted is to show the tags and the values for tagged regular expressions and tagged phrase types.*

EX1005, 11:17-23 (emphasis added); ~~EX1003 ¶154.~~

130. ~~Each~~A POSITA would have understood that each of Richardson's disclosed uses of information extracted by matching regular expressions—i.e., (i) database lookups, (ii) transmission to intent matching module 54, and (iii) preparation of a response comprising tagged regular expressions—is depicted below in Figure 5(b). ~~EX1003 ¶155.)~~

Fig. 5(b)





EX1005, Fig. 5(b).

131. Relatedly, Richardson discloses that “[i]f one of the actions is a response to the writer, the composition of the response message text may include ... phrases and extracted information,” and “for debugging purposes or legal reasons, the input

... phrases and extracted information,” and “for debugging purposes or legal reasons, the input message text, response results and other processing information might be logged for later analysis, reporting or record keeping.” EX1005, 7:41-58. For this additional reason, a POSITA would have understood that Richardson discloses extracting portions of information identified using regular expressions. EX1003 ¶156.

132. Accordingly, in my opinion, Richardson and Madnick teach or suggest all elements of claim 1, which would have been obvious to a POSITA. Id. ¶157.

b. Claim 12

133. Claim 12 recites:

12. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:
at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and
memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the

identifying and extracting operations.

134. In my opinion, Richardson and Madnick render claim 12 obvious for substantially the same reasons as claim 1. ~~EX1003 ¶159. As noted in Ground 1, the~~ The only differences between claims 1 and 12 are that claim 12 recites (i) an apparatus instead of a method; (ii) “at least one processor operative to” perform the claimed steps; and (iii) a final memory limitation. Each of these limitations added by claim 12 would have been obvious to a POSITA in view of Richardson. ~~(iii) a final memory limitation. Id. Each of these limitations in claim 12 would have been obvious to a POSITA over Richardson. Id.~~

First, to the extent the preamble of this claim is limiting, it is my opinion that Richardson discloses it. ~~Id.~~

~~¶160.~~ Richardson is “directed to a method *and apparatus* for automatically recognizing and responding to intent in message text” that performs the functions discussed above for claim 1. EX1005, 1:21-23 (emphasis added). Moreover, Richardson discloses an apparatus—“a computer system”—for performing the process discussed above for claim 1. *Id.*, 3:37-39 (“FIG. 1 is a block diagram of a computer system over which messages are sent to a server for automatic handling

135. according to the present invention.”), 3:47-49 (“FIGS. 5(a)-(b) illustrate *the process of using the apparatus* shown in FIGS. 1-3 to automatically respond to a message.”) (emphasis added); ~~EX1003 ¶160.~~

136. Second, a POSITA would have understood that Richardson discloses at least one processor operative to perform the functions of claims 1 and 12.

~~EX1003 ¶161.~~ In Richardson, “[c]ustomer system 12 sends input message 14 to server 16 *for processing.*” EX1005, 4:40-41 (emphasis added). Server 16 is a “general purpose ~~computer system ... comprising one or more central processing unit....~~” ~~Id., 5:11-15~~ (emphasis added); ~~EX1003~~

computer system ... comprising one or more *central processing unit*....” ¶161.*Id.*,

5:11- 15 (emphasis added). As discussed, server 16 performs the steps of claim 1 (the same steps as claim 12) within server 16’s NLE module 52 comprising regular expression module 60. *Id.*; EX1005, Figs. 1-3, 5:4-6, 5:23-30. Thus, server 16 comprises at least one processor operative to perform the claimed steps. ~~EX1003 ¶161.~~

Third, in my opinion, Richardson teaches or at least suggests claim 12’s memory limitation.

~~*Id.* ¶162.~~ The ’796 patent admits that “memory” broadly includes “any memory devices associated with a processor or CPU, such as, for example, RAM, ROM, a fixed memory device (e.g., hard drive), a removable memory device (e.g., diskette), flash memory, etc.” EX1001, 12:35-39. Richardson’s server 16, which performs the steps of claim 12, comprises “memory (ROM and RAM), and associated peripheral equipment, such as disk drives...,” which are all forms of memory. according to the ’796 patent. EX1005, 5:11-15; ~~EX1003 ¶162.~~

Moreover, Richardson discloses that

137. “[t]he programs used to implement the programming might reside in memory....” EX1005, 5:11-18. Thus, Richardson’s server 16 comprises memory, operatively coupled to the at least one processor (i.e., Richardson’s “one or more central processing unit”). ~~EX1003-¶162;~~ EX1005, 5:11-15.

138. Richardson teaches that server 16’s memory is for storing at least a portion of results associated with the identifying and extracting operations. ~~EX1003-¶163.~~ “In the preferred embodiment, server 16 is ... [an] automated message understanding and

response system.” EX1005, 4:54-56. Specifically, “[s]erver 16 is adapted to take an action based on ... the text of query message,” including “response message 18” (*id.*, 4:44-47), which “may include ... phrases and extracted information” (*id.*, 7:41-44). “A preferred response 130 includes all or part of the original message,” including “tags and the values for tagged regular expressions.” *Id.*, 11:17-23.

Moreover, as I discussed above for claim 1, Richardson teaches other examples in which server 16 uses extracted information to query a database (*id.*, 10:24-30), determine a writer’s intent (*id.*, 10:35-38), and log extracted information for recordkeeping (*id.*, 7:41-58). In each of these examples, a POSITA would have understood that server 16 necessarily stores in memory, at least temporarily, at least a portion of results associated with the identifying and extracting operations—i.e., information extracted using regular expressions—for

purposes of generating a response, querying a database, determining a writer's intent, or maintaining a log. ~~EX1003 ¶163.~~

139. At a minimum, ~~it~~ a POSITA would have ~~been~~ found it obvious to use Richardson's disclosed memory for storing at least a portion of results associated with the identifying and extracting operations. ~~Id. ¶164.~~ A POSITA would have been motivated to store at least a portion of the extracted results in server 16's memory to allow server 16's processor to utilize the results in performing the operations discussed above. Likewise, a POSITA would have been motivated to store the ~~Id.~~

~~Likewise, a POSITA would have been motivated to store the~~ extracted results for later retrieval—e.g., to prepare a follow-up response, to repeat or update a database query, and to log results “for later analysis, reporting or record keeping,” as encouraged by Richardson. EX1005, 7:41-58; ~~EX1003 ¶164.~~

140. Thus, in my opinion, claim 12 would have been obvious. ~~EX1003 ¶165~~
to a POSITA over Richardson and Madnick.

c. Claim 23

141. Claim 23 recites:

23. An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression;
identifying at least a portion of information associated with the at least one regularly identifiable expression; and
extracting the portion of information.

142. In my opinion, Richardson and Madnick render claim 23 obvious for substantially the same reasons as claims 1 and 12. ~~EX1003 ¶167. As noted in Ground 1,~~
~~the~~The only difference between claims 1 and 23 is the preamble: Instead of a

method, claim 23 recites an article of ~~manufacture comprising a machine-readable medium containing one or more programs which when executed implement the same steps as claim 1. *Id.*~~

manufacture comprising a machine-readable medium containing one or more programs which when executed implement the same steps as claim 1.

To the extent the preamble of this claim is limiting, Richardson discloses it. ~~*Id.* ¶168.~~ As I discussed above for claim 12, Richardson is “directed to a method *and apparatus* for automatically recognizing and responding to intent in message text” that performs the steps of claim 1. EX1005, 1:21-23 (emphasis added). Moreover, Richardson discloses “a computer system” for performing the claimed process. *Id.*, 3:37-39.

143. Specifically, “[c]ustomer system 12 sends input message 14 to server 16 for processing” (*id.*, 4:40-41), and server 16 is a “general purpose computer system[] *programmed according to [its] use* and comprising one or more central processing unit, memory (ROM and RAM), and associated peripheral equipment, such as disk drives....” (*id.*, 5:11-15) (emphasis added). Richardson further explains that “[t]he programs used to implement the programming might reside in memory, in one of the storage devices, or a combination of the two.” *Id.*, 5:17-19.

144. Thus, a POSITA would have understood that Richardson discloses an article of manufacture (i.e., server 16) for use in automatically processing an input sequence of data symbols, comprising a machine-readable medium containing one or more programs (i.e., server 16’s memory), which when executed implement the

recited claim steps—the same steps recited in claims 1 and 12. ~~EX1003 ¶169.~~

Similarly, Madnick teaches that its disclosed methods may be performed by “one or more computer-readable programs embodied on or in one or more articles of manufacture,” and “[t]he software programs may be stored on or in one or more

145. articles of manufacture as object code.” EX1006, 16:16-26. A POSITA would have been motivated to store and implement Richardson’s programming on an article of manufacture as taught by Madnick, which also relates to using software to extract information from documents. ~~EX1003 ¶170.~~

146. Thus, in my opinion, claim 23 would have been obvious. ~~EX1003 ¶171~~
to a POSITA over Richardson and Madnick.

d. Claim 25

147. Claim 25 recites:

25. A method of automatically processing an input document,
the method comprising the steps of:
identifying one or more regular expressions in the input
document by comparing the input document to one or more
previously-stored regular expressions to determine if there is
a match between a portion of the input document and at least
one of the previously-stored regular expressions, wherein a
regular expression represents a pattern that is matchable in
accordance with a programming language that supports such
a regular expression; and
identifying at least a portion of information associated with the
at least one regular expression for extraction.

148. In my opinion, Richardson and Madnick render claim 25 obvious for substantially the same reasons as claim 1. ~~EX1003 ¶173. As noted in Ground 1, the~~ The only difference between claims 1 and 25 is that claim 25 omits the final step in claim 1 that requires extracting information and, instead, specifies that regular expressions are identified by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions. ~~Id.~~

~~This~~ A POSITA would have understood that this limitation is disclosed by Richardson. ~~Id. ¶174.~~ First, Richardson's regular expressions are previously stored: the "system isolates the writer's critical words which relate to prototypical statements *stored in the system.*" EX1005, 2:55-57 (emphasis added); ~~EX1003 ¶174.).~~ Specifically, "[e]xtraction rule set 50 is shown comprising a regular expression lexicon 70," such that "the regular expression module 60 receives the input text message and reads from regular expression lexicon 70." EX1005, 5:32-33, 5:38-40. A POSITA would have

149. understood from Richardson that “regular expression lexicon 70” comprising regular expressions is previously stored in order for “regular expression module 60” to “read[] from” it. ~~*Id.*; Fig. 4; EX1003 ¶174.~~

150. Second, Richardson teaches that these previously stored regular expressions are compared to the input document to determine if there is a match.

~~EX1003~~ While reading text, “[r]egular expression module 60 looks for regular expressions

~~¶175. While reading text, “[r]egular expression module 60 looks for regular expressions~~

which are known to regular expression lexicon 70” according to “rules for identifying regular expressions” and “identifies instances of the regular expressions represented in regular expression lexicon 70.” EX1005, 5:51-60.

Specifically, Richardson uses “regular expression list 104,” with an example provided in Table 1. *Id.*, 8:50-66. “Regular expression list 104 is used to isolate and identify character *strings in the message which match the expressions in the list.*” *Id.*, 9:17-19 (emphasis added). This process of comparing and matching previously stored regular expressions is also claimed by Richardson: Claim 1 [of Richardson](#) recites “*comparing* the set of standard lexical tokens to phrase expressions in a collection of lexicon entries, wherein each lexicon entry comprises a phrase expression and an associated phrase type,” and claim 4 [of Richardson](#) specifies that “the phrase type is selected from ... *a regular expression text pattern.*” *Id.*, 11:56-59, 12:8-10 (emphasis added). Similarly, Richardson’s claim 13 recites “*matching* identified tokens to *a database of regular expression patterns.*” *Id.*, 12:51-54 (emphasis

151. added). Madnick also discloses that when using regular expressions, “[f]or each variable to be retrieved in a given state, the state description contains *a pattern to be matched* against the document...” EX1006, 15:1-3 (emphasis added); *see also id.*, 15:5-10; ~~EX1003 ¶176.~~

152. Thus, a POSITA would have understood that Richardson and Madnick ~~teach~~disclose comparing the input document to one or more previously stored

regular expressions to determine if there is a match between a portion of the input document and at least one of the previously stored regular expressions. ~~EX1003~~

153. ~~¶177.~~ Accordingly, in my opinion, claim 25 would have been obvious. ~~Id~~ to a POSITA over Richardson and Madnick.

3. Dependent Claims 2-10 and 13-21

154. It is also my opinion that Richardson and Madnick render obvious each of the following dependent claims.

a. Claims 2 and 13

155. Claims 2 and 13 depend from claims 1 and 12, respectively, and further recite that identifying the regularly identifiable expression “*comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.*”

156. This limitation is substantively identical to the limitation added by claim 25 and, for the same reasons that I discussed ~~above~~ in the preceding section, is taught by Richardson and Madnick. *Supra* §~~v.c~~ VII.B.2.d; ~~EX1003~~ ¶181.

157. Thus, in my opinion, claims 2 and 13 would have been obvious
EX1003 ¶182 to a POSITA over Richardson and Madnick.

b. Claims 3 and 14

158. Claims 3 and 14 depend from claims 1 and 12, respectively, and further recite “*normaliz[ing] the input sequence of data symbols prior to identifying the regularly identifiable expression.*”

159. For context, I note that the ’796 patent provides an example of normalization as “chang[ing]it would be understood by a POSITA:

In step 14, the input data is normalized. This process may include standardizing any characteristics of the input that are irrelevant to, or a hindrance to, information extraction. An example of normalization, for text input, is to change all the letters in the text stream to lower-case and to replace multiple blank spaces between words by a single blank space,”which“. This makes subsequent pattern matching easier.”.

EX1001, 4:~~15~~12-18.

160. Similarly, Richardson discloses that words may be “regularized” and spell-checked ~~before~~prior to identifying regular expressions ~~are identified~~:

Morphology module 66 uses overall lexicon 72 to transform the words into their regularized form. For example, morphology module 66 might convert the word “said” to “say” in an effort to simplify the message. As should be apparent, modules 60, 62, 64, and 66 can operate on the input message non-sequentially. For example, after the spelling module removes a typographical error, the result may match a regular expression which was not earlier matched, in which case

regular expression module 60 would add additional regular expression tags when the message is revisited.

EX1005, 6:9-19; *see also id.*, 9:38-41. These actions, which as described above may occur before regular expressions are identified, comprise normalization and

are similar to the '796 patent's embodiment of changing letters to "make[] subsequent pattern matching easier." EX1001, 4:15-19; ~~EX1003 ¶185~~.

161. Thus, in my opinion, claims 3 and 14 would have been obvious. ~~EX1003 ¶186~~ to a POSITA over Richardson and Madnick.

c. Claims 4 and 15

162. Claims 4 and 15 depend from claims 1 and 12, respectively, and further recite "*identify[ing] one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.*"

163. In my opinion, Richardson discloses this limitation. ~~EX1003 ¶188~~. Specifically, Richardson discloses pre-processing input text to classify data as structured or unstructured:

Prior to the regular expression parse, the message might be pre-processed for structured data. For example, if attachments are included or data is included in the message in a structured format, such as 'HTML' forms data, comma delimited data records, or an attached spreadsheet, that structured data is identified. Preferably, that structured data is flagged such that NLE module 52 does not need to process it.

EX1005, 9:29-36; *see also id.*, 5:40-46 (“the input text is scanned for these data constructs which are not free-form text”), 12:44-50 (claiming step of “[s]eparating free form text in the content from structured text in the content”).

A POSITA would have understood that Richardson’s step for identifying and separating structured data (EX1005, 9:29-36, 12:44-50) comprises identifying classes of data symbols—i.e., structured or unstructured data—and

164. occurs prior to identifying regular expressions. ~~EX1003 ¶189.~~

Likewise, a POSITA would have understood that Richardson's steps ~~for~~of regularizing words and spell-checking discussed above for claims 3 and 14 (EX1005, 6:9-19, 9:38-41; ~~supra~~ §V.C.3.b) also comprise identifying classes of data symbols—i.e., words that may be regularized/simplified or are misspelled—and occur before identifying regular expressions. ~~EX1003~~

165. ~~¶189.~~ Thus, in my opinion, claims 4 and 15 would have been obvious. ~~Id.~~ ¶190 to a POSITA over Richardson and Madnick.

d. Claims 5 and 16

166. Claims 5 and 16 depend from claims 1 and 12, respectively, and further recite: “*wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.*”

In my opinion, Richardson discloses, teaches, or suggests this limitation. ~~EX1003 ¶192.~~

167. ~~For~~As an example, Richardson discloses regular expression list 104, which

includes the following expression to identify a URL: “http://*{ /* }>”. EX1005, 8:62. In this example, the characteristic phrase “http://” typically precedes a particular portion of information—i.e., the rest of the URL—that the regular expression is designed to extract. ~~EX1003 ¶192.~~ Again, this reflects standard and well-known syntax for regular expressions. ~~Id.~~ that would have been well understood to a POSITA.

168. Thus, in my opinion, claims 5 and 16 would have been obvious. ~~Id.~~ ¶193 to a POSITA over Richardson and Madnick.

e. Claims 6 and 17

169. Claims 6 and 17 depend from claims 1 and 12, respectively, and further recite: “*wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.*”

In my opinion, Richardson discloses, teaches, or suggests this limitation. ~~EX1003 ¶195.~~

170. ~~For~~As an example, Richardson discloses regular expression list 104, which includes the following expression to identify an email address: “[a-zA-Z]@*.*”. EX1005, 8:64. In this example, the characteristic phrase “@*.*” typically follows at least a portion of information—i.e., the user name portion of an email address—that the regular expression is designed to extract, ~~which.~~ As with the similar syntax discussed for claims 5 and 16, this would have been well understood to a POSITA. ~~EX1003 ¶195.~~

171. Thus, in my opinion, claims 6 and 17 would have been obvious. ~~Id.~~ ¶196 to a POSITA over Richardson and Madnick.

f. Claims 7 and 18

172. Claims 7 and 18 depend from claims 1 and 12, respectively, and ~~add~~further recite: “*wherein the extracted portion of information is used to take a specified action.*”

In my opinion, Richardson discloses this limitation. ~~EX1003 ¶198.~~

Richardson is directed to “[a]n improved message text understanding and response system,” which uses extracted information to determine “prototypical response actions.” EX1005, 2:53-

63. “The assembled potential actions are evaluated and executed,” and “[e]xamples of potential actions include responding to the writer, giving

173. recommendations for the next step in processing, ... routing the message to a third party for automatic or manual response, or invoking an external system.” *Id.*, 3:9- 15; *see also id.*, Figs. 3, 5(b) (referring to actions based on data extraction and intent mapping), 4:5-7 (“The system determines the customer’s intent, to the extent that it relates to the company’s customer needs, and *takes an action* appropriate for that intent.”), 4:44-47 (“Server 16 is adapted to *take an action* based on an internal knowledge base and the text of query message 14. If the action is a response message 18, that response message 18 is sent to customer system 12.”) (emphases added), 11:62-12:2 (claiming the steps of selecting, generating, and executing actions), 12:37-40 (claiming exemplary actions).

Richardson explains that information extracted using regular expressions is used to determine and take the specified action. ~~EX1003 ¶199~~. For example, “[r]egular expression module 60 looks for regular expressions which are known to regular expression lexicon 70” (*id.*, 5:51-53) and “the input message with all the phrase types identified is passed to an intent matching module (‘IMM’) 54” (*id.*, 6:34-36), which determines “typical cases for which appropriate response actions can be established” (*id.*, 6:42-43). Then, “the message and the intents list is passed to an action mapping module (‘AMM’) 55, which generates a list of proposed actions” (*id.*, 7:18-20) that are processed by “an action resolution module (‘ARM’) 56”

174. (*id.*, 7:35-38), and “[t]he actions taken are ... executed by an action execution module (‘AEM’) 57” (*id.*, 7:45-47). See also *id.*, Fig. 5(b).

175. Thus, in my opinion, claims 7 and 18 would have been obvious ~~EX1003 ¶200~~ to a POSITA over Richardson and Madnick.

g. Claims 8 and 19

176. Claims 8 and 19 depend from claims 1 and 12, respectively, and addfurther recite: “*wherein the extracted portion of information is at least one of visually and audibly presented to the user.*”

177. In my opinion, Richardson discloses this limitation. ~~EX1003 ¶202.~~ As I discussed for ~~claim 1’s~~the final clause of claim 1 (*supra* §~~v.c~~VII.B.2.a.iv), Richardson generates a

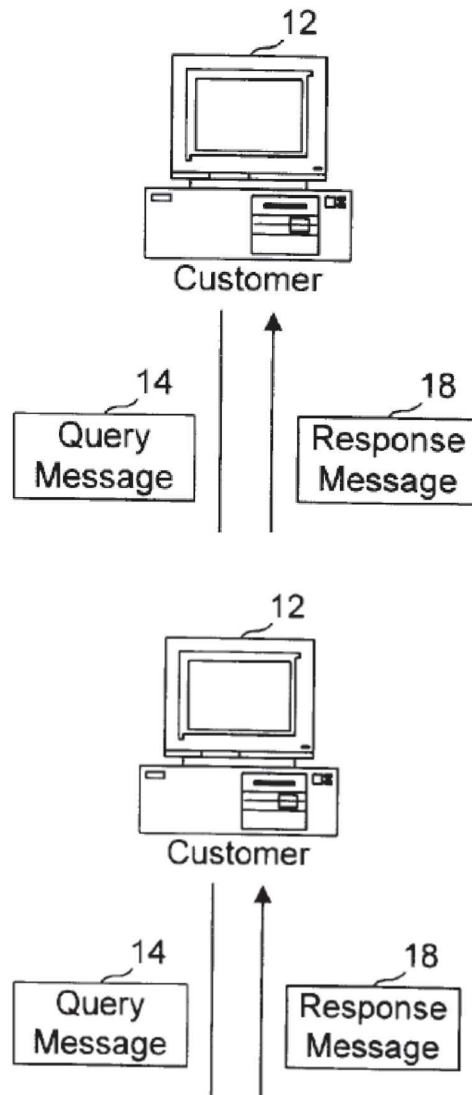
response to the user that includes extracted information identified by regular expressions:

[A] list 122 of actions to take is passed to action execution module (“AEM”) 57, which performs the actions listed on list 122, such as creating a response 130. *A preferred response 130 includes all or part of the original message, ... and an indication of how the message was interpreted. One way to show how the message was interpreted is to show the tags and the values for tagged regular expressions and tagged phrase types.*

EX1005, 11:17-23 (emphasis added); ~~EX1003 ¶202-).~~

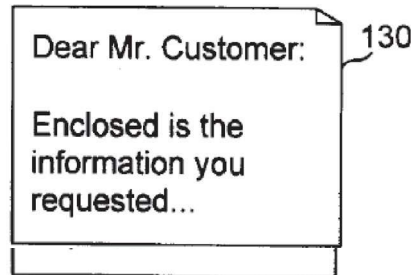
178. ~~This~~A POSITA would have understood that this information is visually presented to the user, who receives the response as text displayed on customer system 12, a “general purpose computer system[] ... comprising one or more ... video display terminals.” ~~EX1005, 5:11-17. As shown~~EX1005, 5:11-17. As shown in Figure 1 of Richardson, partly reproduced below, response message 18 is sent to customer system 12 that comprises such a display (depicted as a CRT monitor):

in Figure 1, response message 18 is sent to customer system 12 that comprises a display (depicted as a CRT monitor):



Id., Fig. 1 (detail).

179. As shown in Figure 5(b), preferred response 130 is a textual message:



Id., Fig. 5(b) (detail); ~~EX1003 ¶204.~~ Accordingly, a POSITA would have understood Richardson discloses that the extracted portion of information is visually presented to the user as part of the textual response sent to customer system

12. ~~EX1003 ¶204.~~

180. At a minimum, since Richardson discloses a video display terminal for customer system 12 that receives a response comprising extracted information, a POSITA would have found it obvious to visually present such information to the user on the provided display terminal. See EX1005, 5:11-17, 11:17-23, Fig. 1.

would have been obvious to visually present such information to the user on the provided display terminal. *Id.* ¶205; EX1005, 5:11-17, 11:17-23, Fig. 1.

181. Thus, in my opinion, claims 8 and 19 would have been obvious. EX1003 ¶206 to a POSITA over Richardson and Madnick.

h. Claims 9 and 20

182. Claims 9 and 20 depend from claims 1 and 12, respectively, and further recite: “*wherein the regularly identifiable expression identifying [step/operation] is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.*”

As I discussed above, while Richardson uses programming in server 16 to match regular expressions and extract information (see EX1005, 5:11-19), Richardson does not expressly identify the specific programming language used. EX1003

183. ¶208. for regular expressions. Supra §VII.B.1. Madnick, however, expressly identifies “PERL” as a “preferred” language because of its “good text-handling capabilities.” EX1006, 16:22-25. Madnick recognizes that “[t]he meaning of PERL regular expressions are well known in the art.” *Id.*, 13:6-7. Madnick further explains perl syntax for regular expressions (*id.*, 13:1-6) and overmore than a dozen examples of regular expressions written in perl (*id.*, col. 12, Tbl. 2). Thus, as shown above in §V.C.VII.B.1, a POSITA would have been motivated to use perl as taught by Madnick in Richardson’s process using regular expressions. EX1003 ¶208.

184. Moreover, consistent with Madnick’s disclosure, the ’796 patent admits that perl has been ~~well~~ known in the art since at least 1996 (EX1001, 3:29-31) and features built-in support for regular expressions (*id.*, 2:26-27).

185. Thus, in my opinion, claims 9 and 20 would have been obvious. ~~EX1003 ¶210~~ to a POSITA over Richardson and Madnick.

i. Claims 10 and 21

186. Claims 10 and 21 depend from claims 1 and 12, respectively, and further recite: “*wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence*

data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.”

187. In my opinion, Richardson discloses this limitation. ~~EX1003 ¶212.~~ Specifically, Richardson’s input sequence of data symbols is an “input text message.” EX1005, 5:34-40; *see also id.*, Figs. 5(a), 5(b) (showing input text 102). Accordingly, a POSITA would have understood that Richardson’s input sequence of data symbols is representative of text data.

188. EX1003 ¶212. Thus, in my opinion, claims 10 and 21 would have been obvious. ~~Id. ¶213.~~

D.C. Ground 3: Claims 11, 22, and 24 Would Have Been Obvious Over Richardson, Madnick, and Groner

189. It is further my opinion that the three other claims of the ’796 patent that are not addressed above—claims 11, 22, and 24—would have been obvious to a POSITS over Richardson, Madnick, and Groner. Below, first, I discuss why a POSITA would have further combined the Richardson-Madnick combination I addressed above with Groner. Then, I address how the Richardson-Madnick-Groner combination meets all limitations of these claims.

1. Richardson-Madnick-Groner Combination

190. As discussed above, Richardson expressly relates to “automatically recognizing and responding to intent in message text, particularly as used in a large-scale automatic message response system....” EX1005, 1:21-26.

Specifically, Richardson describes a method and system for responding to

~~system for responding to~~ “electronic mail” (e.g., *id.*, 3:19, 4:23), including in the context of “a company’s customer service department” (*id.*, 4:2-5). See also *id.*, Figs. 5(a), 5(b) (depicting customer emails). Although Richardson acknowledges the prevalence of “other messaging media, such as telephone calls” (*id.*, 1:32-34), Richardson’s method and system are not expressly described in terms of any messages in forms other than text. ~~EX1003 ¶215.~~ (e.g., emails). Thus, in my opinion, a POSITA would have been motivated to seek techniques for implementing Richardson’s method and system on these “other messaging media, such as telephone calls.” that were well known. *Id.*

191. This would have led a POSITA to search for and locate Groner. In the same field of endeavor as Richardson, Groner explicitly ~~describes~~ relates to “speech recognition systems as applied to voice and electronic message mailing, and particularly to a system and method for converting speech to a text message suitable for sending as an e-mail message.” EX1007, 1:6-10. Groner discloses a method and system in which “[a]n audio message from a caller for a recipient is received,” “[a] text message file is generated from the audio message from the caller,” and “[t]he text message file is sent to the recipient at the recipient’s e-mail address.” *Id.*, 2:63-67.

192. A POSITA would have been motivated to combine Richardson- and Madnick with Groner to allow Richardson’s input messages (e.g., from customers)

to be received over the phone via voice mail and then converted to text for
~~identifying and extracting information with regular expressions. EX1003 ¶217. Groner expressly~~

purposes of identifying and extracting information with regular expressions.

Groner expressly motivates this combination, explaining that “[v]oice mail messages ... are convenient because telephones are ubiquitous and inexpensive.”

EX1007, 2:17-20; ~~EX1003 ¶217.~~ Moreover, by converting voice mail to text, “the voice messages can be searched for particular content.” ~~EX1007, Id.~~, 3:37-39.

193. Richardson acknowledges that customers use telephones to communicate since one of its exemplary regular expressions is for identifying telephone numbers. in customer emails. EX1005, 5:55, 8:63, 9:26-27. While Richardson highlights the “increased popularity of electronic messaging,” it acknowledges “other messaging media, such as telephone calls.” *Id.*, 1:32-34. Groner’s ~~voicemail~~-method of converting voice mail to ~~email-conversion~~ fits logically with Richardson’s exemplary embodiment of a company receiving customer inquiries because customers may use both email and phone calls to communicate. ~~EX1003 ¶218.~~

194. In combining Richardson-Madnick with Groner, a POSITA would have had a reasonable expectation of success. ~~Id. ¶219.~~ Because Groner converts voice messages to emails, Richardson’s method of processing emails requires no modification. ~~Id. Combining~~ A POSITA would have understood that combining Richardson-~~Madnick~~ with Groner simply allows customers to contact the company by phone in addition to the existing ability to contact the company by

email. ~~Id.~~ Once the voice

message is transcribed to text and sent via email as described in Groner, it is processed in the same manner as any other email as described in Richardson. ~~Id.~~

195. Implementing Groner in the context of Richardson-Madnick would have been routine to a POSITA. ~~Id.~~ ¶220. Groner notes that “[t]he accuracy of speech recognition software has improved,” with several commercial embodiments available, and expressly discloses a computer system with a “voice-to-electronic mail system,” including “a speech recognition procedure.” *Id.*, 2:38-49, 5:5-9, 6:6-8. Moreover, as with Richardson-and Madnick, Groner’s invention “is implemented as a computer program product that includes a computer program mechanism embedded in a computer readable storage medium.” *Id.*, 16:35-39. Thus, in my opinion, it would have been routine for a POSITA to implement Groner’s software in Richardson’s computer hardware, i.e., server 16. ~~EX1003~~ ¶220.

2. Claims 11 and 22

196. Claims 11 and 22 depend from claims 1 and 12, respectively, and additionally recite: “*wherein the input sequence of data symbols is representative of a voice mail message.*”

197. In my opinion, the Richardson-Madnick-Groner combination discussed above meets this limitation. ~~EX1003~~ ¶222. Groner teaches that “[a]n audio message from a caller for a recipient is received,” “[a] text message file is generated from the audio message from the caller,” and “[t]he text message file is sent to the

recipient at the recipient's e-mail address." EX1007, 2:63-67. When combined with Richardson-Madnick, the recipient is a company's customer service department, which processes the input sequence of data symbols in the transcribed email in the same manner discussed in claim 1. *Supra* §VII.C.1. Thus, in the Richardson-Madnick-Groner combination, the input sequence of data symbols is representative of a voice mail message.

~~in the transcribed email as discussed for claim 1. *Supra* §V.C.2.a; EX1003 ¶222.~~
~~Thus, in Richardson Madnick Groner, the input sequence of data symbols is~~
~~representative of a voice mail message. EX1003 ¶222.~~

198. Accordingly, in my opinion, claims 11 and 22 would have been obvious. ~~*Id.* ¶223~~ to a POSITA over Richardson, Madnick, and Groner.

3. Claim 24

~~Claim 24 would have been obvious for substantially the same reasons~~
~~discussed above for claims 1 and 12 in Ground 2. *Supra* §§V.C.2.a-b; EX1003~~

199. ~~¶225.~~ The only difference between claims Claim 24 recites as follows:

~~24 and 12 is that claim 24 adds (i) “.~~ Apparatus for
automatically processing an input sequence of data symbols, the
apparatus comprising:

a data capture device for obtaining the input sequence of data
symbols²²;

at least one processor, operatively coupled to the data capture
device, operative to: (i) identify at least one regularly
identifiable expression in the input sequence of data
symbols. wherein the at least one regularly identifiable
expression represents a pattern that is matchable in
accordance with a programming language that supports such
a regularly identifiable expression; (ii) identify at least a
portion of information associated with the at least one
regular expression; and (iii) extract the portion of

| [information;](#)

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and (ii) “extracting operations; and
a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user.” ~~EX1001, claim 24. Both limitations are met by Richardson-Madnick-Groner. EX1003 ¶225.~~

200. In my opinion, claim 24 would have been obvious for substantially the same reasons that I discussed above for claims 1 and 12 in Ground 2. Supra §§VII.B.2.a-b. The only difference between claims 24 and 12 is that claim 24 adds (i) “a data capture device for obtaining the input sequence of data symbols”; and (ii) “a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user.” In my opinion, both limitations are met by the Richardson-Madnick-Groner combination.

201. First, as to the “data capture device,” the ’796 patent explains that a “data capture device may be a speech processing device capable of taking an audio stream recorded in accordance with a voice mail system and converting or transcribing the audio stream to a discrete data or symbol sequence.” EX1001, 11:53-57. A POSITA would have understood that Groner discloses this embodiment. ~~EX1003 ¶226.~~ Specifically, Groner discloses “voice-to-electronic mail system 30,” which “automatically converts a spoken message to a text message which is e-mailed to a recipient.” EX1007, 5:5-9. The system comprises “a

speech

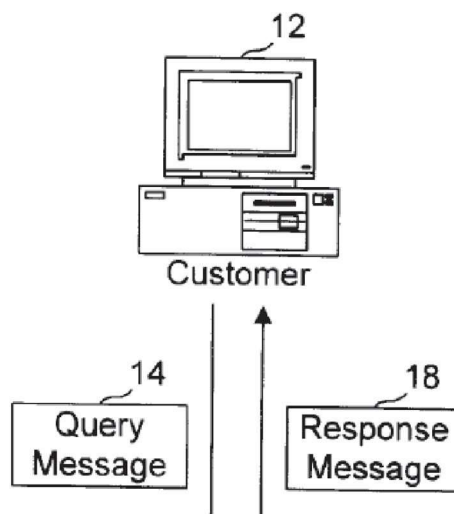
9. The system comprises “a speech

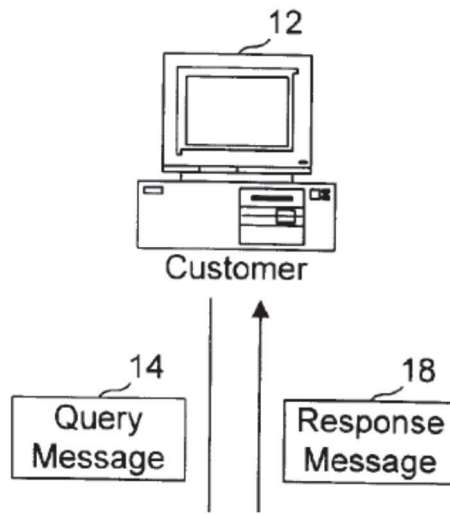
| recognition procedure 116 that receives audio

speech, identifies the audio speech and generates a text file 118 corresponding to the identified audio speech.” *Id.*, 6:6-8. Thus, Groner’s voice-to-electronic mail system 30 is a data capture device. ~~EX1003 ¶226.~~

Second, as to the “data output device,” the ’796 patent explains that an embodiment of a “data output device [] generally presents the data extracted from the input data stream to a user” and “may be a display for visually presenting the extracted data to the user.” EX1001, 12:52-53, 12:62-63. ~~In~~ Likewise, in Richardson, messages are received and sent to customer system 12, a “general purpose computer system[]

202. ... comprising one or more ... video display terminals.” EX1005, 5:11-17. For example, as shown in Figure 1, which is partly reproduced below, response message 18 is sent to customer system 12 comprising a display (depicted as a CRT monitor):





203.

Id., Fig. 1.

As [I](#) discussed [above](#) for claim 1, “the response message text may include ... phrases and extracted information” (*id.*, 7:42-44), including “tags and the values for tagged regular expressions (*id.*, 11:17-23). See [also](#) *id.*, Fig. 5(b) (preferred

204. response 130 to customer); EX1003 ¶229.) A POSITA would have understood that the response generated by Richardson, which includes an extracted portion of information, would be displayed to the customer on a data output device like such as the display of customer system 12. EX1003 ¶229.

205. At a minimum, it a POSITA would have been found it obvious to use the display of Richardson's customer system 12 in Richardson to output response messages comprising extracted information given Richardson's goal of automating responses to customer inquiries. Id. ¶230. Thus, claim 24 would have been obvious. Id to a POSITA over Richardson, Madnick, and Groner.

E.D. No Secondary Considerations Exist

No I am unaware of any secondary indicia considerations of nonobviousness exist. Id. ¶231. Petitioner reserves, and I understand no such considerations were alleged during prosecution. I reserve the right to respond to rebut any assertion of alleged secondary indicia by considerations that Patent Owner.

IV. The Parallel District Court Proceeding Does Not Warrant a Discretionary Denial

Under the Director's interim procedure for evaluating requests for discretionary denial under 35 U.S.C. §314(a), regardless of any factors that otherwise apply, "the PTAB will not discretionarily deny institution of an IPR or PGR in view of parallel district court litigation where a petitioner stipulates not to pursue in a parallel district court proceeding the same grounds as in the petition or its expert declarant may assert during any grounds that could have reasonably been

raised in the petition.” *Interim Procedure for Discretionary Denials in AIA Post-Grant Proceedings with Parallel District Court Litigation*, 7 (USPTO Dir. June 21, 2022).

~~206. Petitioner stipulates that, should the Board institute~~instituted *inter partes* review, Petitioner will not pursue in parallel district court proceedings the same grounds as in the Petition or any grounds that could have reasonably been raised in the petition. Thus, the Board should not exercise discretion to deny institution under §314(a).

~~V. —~~ **Mandatory Notices**

~~A. —~~ **Real Parties in Interest**

~~Petitioner is the real party in interest. No other party had access to or control over the present Petition, and no other party funded or participated in preparation of the present Petition.~~

~~B. —~~ **Related Proceedings**

~~No previous IPR or post-grant proceeding has been filed challenging the '796 patent. The '796 patent has been asserted by Patent Owner in the following district court litigations: *Taasera Licensing LLC v. Trend Micro Inc.*, No. 2-21-cv-00441 (E.D. Tex.); *Taasera Licensing LLC v. Palo Alto Networks, Inc.*, No. 2-22-cv-00062 (E.D. Tex.); *Taasera Licensing LLC v. Check Point Software Techs. Ltd.*, No. 2-22-cv-00063 (E.D. Tex.); *Trend Micro Inc. v. Taasera Licensing LLC*, No. 3-22-cv-00477 (N.D. Tex.); *Trend Micro, Inc. v. Taasera Licensing LLC*, No. 3-22-cv-00518 (N.D. Tex.); *Palo Alto Networks, Inc. v. Taasera Licensing LLC*, No. 1-22-cv-02306 (S.D.N.Y.); *In re: Taasera Licensing LLC Patent Litig.*, No. 2-22-md-03042 (E.D. Tex.); *Trend Micro Inc v. Taasera Licensing LLC*, No. 2-22-cv-~~

~~00303 (E.D. Tex.); *Palo Alto Networks, Inc. v. Taasera Licensing LLC*, No. 2:22-
cv-00314 (E.D. Tex.).~~

V. Conclusion

207. In sum, for the reasons I have discussed above, it is my opinion that claims 1-25 of the '796 patent are unpatentable because they would have been obvious to a POSITA over the prior art references cited above.

208. I understand that this declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also understand that I may be subject to cross-examination concerning this declaration, and I will appear for cross-examination, if required of me, during the time allotted for cross-examination.

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