

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

CNET NETWORKS, INC.,

Plaintiff,

No. C 06-5378 MHP

v.

ETILIZE, INC.,

Defendant.

MEMORANDUM & ORDER
Defendant's Motion for Summary
Judgment on Noninfringement

Plaintiff CNET Networks, Inc. ("CNET") filed this action against Etilize, Inc. ("Etilize") alleging infringement of two patents that teach a process of compiling information about various consumer products into a database. Now before the court is defendant Etilize's motion for summary judgment on noninfringement of the asserted claims of the two patents, United States Patent Nos. 6,714,933 ("the '933 Patent") and 7,082,426 ("the '426 Patent"). Having fully considered the parties' arguments and submissions¹ and for the reasons set forth below, the court enters the following memorandum and order.

BACKGROUND

I. **The Patented Inventions**

This infringement action relates to patents claiming methods and processes that aggregate content for online purchasing and cataloging systems. The '933 Patent, issued on March 30, 2004, and the '426 Patent, issued on July 25, 2006, disclose methods of aggregating product information from a plurality of sources using crawlers, computational linguistics and software.

1 A. The ‘933 Patent

2 The ‘933 Patent, titled “Content Aggregation Method and Apparatus For On-Line
3 Purchasing,” claims methods of gathering information about various products from multiple sources
4 for storage in a product database. The two claims at issue for the ‘933 Patent are Claims 1 and 15.

5 Claim 1 claims:

6 A method of aggregating product information for use in a product database
7 including various products arranged in product categories, the product information
8 being collected from a plurality of sources in a networked computer environment
9 regarding products of a product category comprising the steps of:

10 generating a crawler from a server interconnected to the network computer
11 environment to visit the plurality of sources;

12 gathering product phrase information from each of the plurality of sources
13 via said crawler; and

14 determining whether at least one phrase of said product phrase information
15 is a product characteristic associated with a product category;

16 wherein said crawler utilizes computational linguistics to gather said
17 product phrase information which includes a phrase and at least one
18 characteristic of said phrase.

19 ‘933 Patent at 18:49–65. Claim 15 depends upon the method of Claim 1.

20 Specifically, a crawler generated from a server interconnected to a network computer
21 environment gathers pertinent phrase information from a plurality of sources. In other words, a
22 crawler—a software program that visits websites and has the ability to identify and gather
23 information from these sites—is launched from a computer connected to the internet. This crawler
24 scours different websites on the internet to gather pertinent information. Pertinent information is
25 determined through the use of computational linguistics—a field of statistical or rule based modeling
26 of natural language that uses computational analysis. The crawler then determines whether at least
27 one phrase of the phrase information gathered is a product characteristic associated with a product
28 category.

29 B. The ‘426 Patent

30 The ‘426 Patent, titled “Content Aggregation Method and Apparatus For An On-Line
31 Product Catalog,” is a continuation-in-part of the ‘933 Patent. It populates a catalog by categorizing

1 and storing into the catalog product information from multiple web pages. The '426 Patent claims
2 methods of processing disparate product information records from various sources into one or more
3 groups. Determining the group in which to place a particular record depends on which product
4 information records are likely to correspond to the same product. Each group, which corresponds to
5 a particular product, is given a unique identifier. This identifier is then compared to categories in a
6 taxonomy to determine a category for that particular product in the taxonomy. Finally, product
7 attributes are determined for each categorized product based on the earlier collected product
8 information records for that product.

9 There are eight claims at issue for the '426 Patent—Claims 1, 14, 16, 20, 23, 24, 39 and 52.
10 Of these only Claims 1, 39 and 52 are independent while the rest are dependent.

11 1. Claim 1

12 Claim 1 of the '426 Patent claims as follows:

13 A method of creating a product catalog stored on computer readable media by
14 aggregating product information from a plurality of product information sources
15 having disparate formats for product information and storing the information in a
16 taxonomy, said method comprising:

17 processing plural product information records from the product information
18 sources into one or more groups based on which product information
19 records are likely to correspond to the same product;

20 correlating a unique product ID corresponding to the product associated
21 with each of said groups to identify the product;

22 electronically comparing each identified product to categories of a
23 taxonomy to determine a category for the identified products in the
24 taxonomy; and

25 electronically parsing the product information records corresponding to each
26 group to electronically determine attributes for each categorized product
27 based on the product information records;

28 electronically generating product specifications based on the determined
attributes; and

storing the product specification in the corresponding determined categories
of the taxonomy.

'426 Patent at 36:22–45. Claims 14, 16, 20, 23 and 24 depend upon Claim 1.

2. Claim 39

Claim 39 of the '426 Patent is substantially similar to Claim 1. However, Claim 39 does not

1 store the information comprising the product catalog in a taxonomy. Furthermore, Claim 39 repeats
2 the processing and correlating steps after performing the comparing step to revise the groups in
3 which the product information records belong. See ‘426 Patent at 39:66–40:23.²

4 3. Claim 52

5 Claim 52 of the ‘426 Patent claims as follows:

6 A method of aggregating product information from a plurality of product
7 information sources in a networked computer environment comprising the steps of:

8 generating a crawler from a server interconnected to the network computer
9 environment to visit the plurality of sources;

10 gathering product phrase information and characteristics of said product
11 phrase information from each of the plurality of sources via said crawler;

12 grouping said product phrase information based on which product phrase
13 information are likely to correspond to the same product and based on the
14 characteristics of said product phrase information;

15 electronically parsing said grouped product phrase information to determine
16 attributes for each product based on at least one of the product phrase
17 information and the characteristics of said product phrase information; and

18 creating a catalog of products based on the determined attributes.

19 ‘426 Patent at 41:36–56.

20 C. Claim Terms

21 The court uses the following definitions for claim terms that are implicated by the instant
22 motion:

23 Source	A webpage or other document that may be defined by a URL, or text, graphics, or links within such webpage or other document.
24 Computational Linguistics	A cross-disciplinary field of modeling of language utilizing computational analysis to process language data.
25 Phrase	A string of characters, such as an alpha-numeric character string or strings present in a source.
26 Crawler	A software program or programs which visit and search sources of content on a networked computer environment; have the capability to identify and gather information from the sources; and can include bots, robots, automated site searchers, and the like.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Phrase Characteristics	Some attribute of a phrase that can be used to distinguish it from other phrases, such as the frequency, location, font size, font style, font case, font effects, and font color as well as the frequency of collocation (phrases immediately next to each other) and co-occurrence of phrases (phrases within a predetermined number of words of each other) includes the co-occurrence of phrases, which is when a particular word is within a number of words of another (e.g., “weight” and “lbs.”).
------------------------	---

See generally Docket Nos. 66, 82.

II. The Accused Product

The accused product, SpeX, is created by Etilize through the use of a pair of software tools called aQuire and Xtract. Etilize uses these tools in situations where manufacturers store information about their products in a consistent manner on their website. These tools help extract information from the website in an automated fashion. Hameed Dec. ¶ 1. aQuire is a semi-automated search script used to download entire web pages describing a particular product’s specifications. Id. ¶ 11. Xtract is then used to extract data from these downloaded web pages for entry into Etilize’s templates. Id. ¶ 12.

A. aQuire

The aQuire tool fetches web pages by using predefined URLs and simple searching patterns. Id. ¶ 11. aQuire downloads all of the web pages from a specified URL directory by sequentially accessing the web pages associated with each product. It does so by sequentially substituting different model numbers and their associated product categories into the URL template. Id. For example, if the product category is “laptop computers” made by Dell and Dell has five separate models associated with this product category, then each such model number and the product category “laptop computers” would be substituted into the URL template. Id. For instance, aQuire would first download www.dell.com/laptop_comptuer/model#1, followed by www.dell.com/laption_computer/model#2, and so on. This process would repeat until all web pages for the models known to aQuire were downloaded by aQuire. Id. aQuire does not look for unknown

1 products; it is configured by humans to fetch web pages associated with predefined products if a
2 manufacturer’s website is consistently structured. Id. Specifically, aQuire simply stores “copies of
3 the pages that it accesses on an Etilize Pakistan server to facilitate data extraction at a later time.”

4 Id.

5
6 B. Xtract

7 After aQuire downloads and stores copies of certain web pages on the Etilize server, Xtract is
8 used to both “obtain raw attribute/value pairs (*e.g.* attribute: hard drive capacity; value: 40GB)”
9 from these downloaded websites and to store these gathered pairs into the relevant product
10 categories in Etilize Pakistan’s templates. Id. ¶ 12. In order to gather and store these attribute/value
11 pairs, Xtract utilizes predefined expression patterns to parse the downloaded web pages. Id. The
12 Xtract software tool is semi-automated and researchers must specify, for each product attribute: 1)
13 where to extract the information from the web page; 2) what information to extract; and 3) where to
14 put it in the Etilize template. Id.

15
16 LEGAL STANDARD

17 I. Summary Judgment

18 As in any other civil action, summary judgment is proper in a patent infringement action
19 when the pleadings, discovery and affidavits show that there is “no genuine issue as to any material
20 fact and that the moving party is entitled to judgment as a matter of law.” Fed. R. Civ. P. 56(c); see
21 also Southwall Techs., Inc. v. Cardinal IG Co., 54 F.3d 1570, 1575 (Fed. Cir.), cert. denied, 516 U.S.
22 987 (1995). Material facts are those which may affect the outcome of the case. Anderson v. Liberty
23 Lobby, Inc., 477 U.S. 242, 248 (1986). A dispute as to a material fact is genuine if there is sufficient
24 evidence for a reasonable jury to return a verdict in favor of the nonmoving party. Id.

25 The party moving for summary judgment bears the burden of identifying those portions of
26 the pleadings, discovery and affidavits that demonstrate the absence of a genuine issue of material
27 fact. Celotex Corp. v. Cattrett, 477 U.S. 317, 323 (1986). On an issue for which the opposing party
28

1 will have the burden of proof at trial, the moving party need only point out “that there is an absence
2 of evidence to support the nonmoving party’s case.” Id. at 325; Crown Operations Int’l, Ltd. v.
3 Solutia, Inc., 289 F.3d 1367, 1377 (Fed. Cir. 2002). On the other hand, where the moving party
4 bears the burden of proof on an issue, it must submit evidence sufficient to establish that no
5 reasonable jury could find against it on that issue at trial. See Frank’s Casing Crew & Rental Tools,
6 Inc. v. Weatherford Int’l, Inc., 389 F.3d 1370, 1376 (Fed. Cir. 2004); Gart v. Logitech, Inc., 254
7 F.3d 1334, 1339 (Fed. Cir. 2001), cert. denied, 534 U.S. 1114 (2002).

8 Once the moving party meets its initial burden, the nonmoving party must go beyond the
9 pleadings and, by its own affidavits or discovery, “set forth specific facts showing that there is a
10 genuine issue for trial.” Fed. R. Civ. P. 56(e). Mere allegations or denials do not defeat a moving
11 party’s allegations. Id.; Gasaway v. Nw. Mut. Life Ins. Co., 26 F.3d 957, 960 (9th Cir. 1994). The
12 court may not make credibility determinations, and inferences to be drawn from the facts must be
13 viewed in the light most favorable to the party opposing the motion. Masson v. New Yorker
14 Magazine, 501 U.S. 496, 520 (1991); Anderson, 477 U.S. at 249.

15
16 II. Patent Infringement

17 Under the Patent Act, 35 U.S.C. section 271, liability for patent infringement may be
18 imposed on any person who without permission of the patentee “makes, uses, offers to sell, or sells
19 any patented invention[] within the United States or imports into the United States any patented
20 invention during the term of the patent therefor.” The rights granted to the patentee are defined by
21 the patent’s claims. Markman v. Westview Instruments, Inc., 517 U.S. 370, 373 (1996).

22 In determining whether an allegedly infringing device falls within the scope of the claims, a
23 two-step process is used: first, the court must determine as a matter of law the meaning of the
24 particular claim or claims at issue; and second, it must consider whether the accused product
25 infringes one or more of the properly construed claims. Id. at 384; Allen Eng’g Corp. v. Bartell
26 Indus., Inc., 299 F.3d 1336, 1344 (Fed. Cir. 2002). The second inquiry is a question of fact,
27 although summary judgment of infringement or noninfringement may nonetheless be appropriate
28

1 when no genuine dispute of material fact exists. Irdeto Access, Inc. v. Echostar Satellite Corp., 383
2 F.3d 1295, 1299 (Fed. Cir. 2004) (quoting Bai v. L & L Wings, Inc., 160 F.3d 1350, 1353 (Fed. Cir.
3 1998)).

4 The patentee bears the burden of proving infringement by a preponderance of the evidence.
5 Laitram Corp. v. Rexnord, Inc., 939 F.2d 1533, 1535 (Fed. Cir. 1991). This burden can be met by
6 showing that the patent is infringed either literally or under the doctrine of equivalents. See Linear
7 Tech. Corp. v. Impala Linear Corp., 379 F.3d 1311, 1318 (Fed. Cir. 2004). To support a finding of
8 literal infringement, the patentee must establish that “every limitation recited in the claim appears in
9 the accused product, i.e., the properly construed claim reads on the accused product exactly.”
10 Jeneric/Pentron, Inc. v. Dillon Co., 205 F.3d 1377, 1382 (Fed. Cir. 2000) (citing Amhil Enters. Ltd.
11 v. Wawa, Inc., 81 F.3d 1554, 1562 (Fed. Cir. 1996)).

12 Alternatively, where one or more elements of the claim are not literally present in the
13 allegedly infringing product or process, infringement may nonetheless be found under the doctrine
14 of equivalents if the differences between the accused device and the patented invention are
15 “insubstantial.” Honeywell Int’l, Inc. v. Hamilton Sundstrand Corp., 370 F.3d 1131, 1139 (Fed. Cir.
16 2004) (quoting Eagle Comtronics, Inc. v. Arrow Communication Labs., Inc., 305 F.3d 1303, 1315
17 (Fed. Cir. 2002)), cert. denied, 537 U.S. 1172 (2003). As with literal infringement, the inquiry into
18 whether infringement may be found under the doctrine of equivalents requires an element-by-
19 element comparison of the patented invention to the accused device. Warner-Jenkinson Co. v.
20 Hilton Davis Chem. Co., 520 U.S. 17, 40 (1997). Consequently, in applying the doctrine, the court
21 must consider whether the accused device “contain[s] elements that are either identical or equivalent
22 to each claimed element of the patented invention.” Id.; EMI Group N. Am., Inc. v. Intel Corp., 157
23 F.3d 887, 896 (Fed. Cir. 1998), cert. denied, 526 U.S. 1112 (1999).

24 Under the classic formulation of the doctrine of equivalents set forth in Graver Tank & Mfg.
25 Co. v. Linde Air Prods. Co., 339 U.S. 605, 608 (1950), a feature of the accused device is
26 “equivalent” to an element of the claimed invention if it performs substantially the same function in
27 substantially the same way to achieve substantially the same result. See also Schoell v. Regal Mar.
28

1 Indus., Inc., 247 F.3d 1202, 1209–10 (Fed. Cir. 2001). However, as the Supreme Court
2 subsequently acknowledged in Warner-Jenkinson, this particular “linguistic framework” may not be
3 appropriate in every case. 520 U.S. at 39–40. Rather, the Court observed that “[a]n analysis of the
4 role played by each element in the context of the specific patent claim [must] inform the inquiry as
5 to whether a substitute element matches the function, way, and result of the claimed element, or
6 whether the substitute element plays a role substantially different from the claimed element.” Id. at
7 40. A number of other considerations may also be relevant in determining the range of equivalents
8 to which the claimed invention is entitled, including the prosecution history of the patent-in-suit, the
9 pioneer status of the invention (or lack thereof), and the limitations on patentability of the allegedly
10 equivalent device that would have been imposed by the existing prior art at the time that the patent
11 application was filed. Intel Corp. v. Int’l Trade Comm’n, 946 F.2d 821, 842 (Fed. Cir. 1991); see
12 also K-2 Corp. v. Salomon S.A., 191 F.3d 1356, 1366–68 (Fed. Cir. 1999).

13
14 DISCUSSION

15 Etilize argues that multiple limitations in the claims of the patents-in-suit are not practiced by
16 its devices. Before reaching the merits, the court discusses a threshold matter with respect to the
17 manual processes employed by Etilize.

18
19 I. Manual Processes

20 The court notes that Etilize’s manual collection of product information for input into the
21 taxonomy, which it utilizes a majority of the time, cannot infringe upon either the ‘933 Patent or the
22 ‘426 Patent. During patent prosecution, in order to distinguish its ‘426 Patent application from
23 Blutinger, a prior art reference, CNET disavowed: 1) the manual determination of attributes; 2) the
24 manual comparison of products to categories of the taxonomy; and 3) the manual generation of
25 product specifications based on the determined attributes. Khaliq Dec., Exh. G at 28. CNET stated
26 that “in contrast [to Blutinger] the present invention parses the product information records
27 corresponding to each group electronically to determine the attributes;” makes comparisons
28 “between the [sic] each identified product to categories of a taxonomy . . . electronically, not

1 manually;” and generates product specifications based on the determined attributes electronically.
2 Id. Further, during patent prosecution, CNET also disavowed the manual collection of desired
3 product information for input into the database in order to distinguish its ‘933 Patent application
4 from Blutinger. CNET stated that “Blutinger does not disclose or suggest implementing [a system
5 utilizing a crawler] to extract information from a website or a URL, but rather, discloses a
6 centralized master catalog that is essentially manually generated.” Id. at 30.

7 Etilize, similar to Blutinger, utilizes a team of researchers who “typically go to a
8 manufacturer’s website, . . . gather the desired information, and fill the gathered information into the
9 relevant fields of the template.” Hameed Dec. ¶ 11. Further, Etilize utilizes a template based system
10 for categorizing, comparing and generating product specifications and the creation of this template is
11 a manual process. Id. ¶ 9. Accordingly, Etilize’s use of manual methods to perform the comparing,
12 determining, generating and gathering functions cannot infringe either of CNET’s patents.

13 The court turns now to CNET’s patent infringement claim based on Etilize’s use of software
14 tools to extract information from websites.

15
16 II. Infringement of the ‘933 Patent

17 Etilize argues its software tools do not infringe the ‘933 Patent because they do not (1) use a
18 crawler (2) generated from a server (3) that visits a plurality of sources and (4) uses computational
19 linguistics. The court discusses each in turn.

20
21 A. Crawler

22 A crawler is “a software program or programs which visit and search sources of content on a
23 networked computer environment; have the capability to identify and gather info from the sources;
24 and can include bots, robots, automated site searchers and the like.” Docket No. 82 at 6. Rather
25 than arguing that it does not employ software that falls within the construction of crawler provided
26 by the court, Etilize focuses on the fact that because it does not use a server to deploy a crawler, and
27 because it does not utilize a program which visits a “plurality of sources,” it cannot be using a
28 crawler. Both the “server” and “plurality of sources” issues are discussed below. However, to the

1 extent that Etilize argues that it does not utilize a crawler, the court rejects this argument. aQuire is
2 a “software program” that visits web pages or “sources” of content in a “networked computer
3 environment” (i.e., the internet). Further, Xtract is a software tool that has the ability to “identify
4 and gather” information from these web pages. Accordingly, the combination of aQuire and Xtract
5 constitutes use of a crawler by Etilize.

6
7 B. Generated from a Server Interconnected to a Network Computer Environment

8 Etilize argues that aQuire is deployed by human operators from individual client computers
9 connected to the internet, which it contends are not servers as required by Claim 1 of the ‘933 Patent
10 and Claim 52 of the ‘426 Patent. Thus, Etilize is essentially making two arguments for the
11 proposition that it does not generate a crawler from a server: 1) the crawler is not automatically
12 generated but instead is deployed by human operators; and 2) the crawler is deployed from an
13 individual client computer which does not constitute a server.

14 Etilize’s first argument has already been rejected by this court in its claim construction order.
15 In rejecting Etilize’s argument that a crawler must operate without human intervention, the court
16 stated that “crawlers are not intended or claimed as software which operate perpetually, without any
17 human intervention or instruction. Neither patent disclaims human initiation of the crawler search.”
18 Docket No. 82 at 12. Indeed, neither patent speaks to who deploys the crawler. Accordingly, a
19 crawler deployed by a human operator can meet the “generating a crawler from a server” limitation
20 at issue here.

21 The court now turns to Etilize’s second argument that it does not generate a crawler from a
22 server. The preferred embodiments of both the ‘933 and ‘426 patents state that a server “refers to
23 any type of computing device . . . such as a personal computer, a portable computer . . . a hand held
24 device, a wireless phone, or any combination of such devices. The various clients and servers can be
25 a *single computer* at a *single location*.” ‘933 Patent at 10:24–28 (emphases added); see also ‘426
26 Patent at 8:34–38. Accordingly, because a single computer can constitute a server, and Etilize
27 utilizes a crawler that is deployed from individual client computers connected to the internet, Etilize
28 practices this limitation found in Claim 1 of the ‘933 Patent and Claim 52 of the ‘426 Patent.

1 Alternatively, this limitation is met under the doctrine of equivalents even if a “computer” connected
2 to the internet, as opposed to a “server,” generates the crawler.

3
4 C. Plurality of Sources

5 A “source” is “a webpage or other document that may be defined by a URL, or text,
6 graphics, or links within such webpage or other document.” Docket No. 66, Exh. A at 1. This
7 agreed upon definition is not surprising as the preferred embodiments of both the ‘933 and ‘426
8 patents state that the “crawler may crawl through the plurality of Web pages linked to the home Web
9 page to gather product phrase information.” ‘933 Patent at 12:55–58; ‘426 Patent at 11:11–14.

10 Etilize now contends that the term “plurality of sources” should be construed as referring to
11 web pages hosted by separate manufacturers or merchants. Thus, it is conflating the definition of
12 “source” with that of an internet domain. However, as is clear from the parties’ agreed construction
13 of “source,” a plurality of sources can be a plurality of web pages and is not limited to a plurality of
14 separately hosted websites. The plain language of the construction states that a source is a “webpage
15 . . . that may be defined by a URL.” There is no argument that aQuire visits a multitude of web
16 pages, each defined by a separate URL. Consequently, this limitation is practiced by the software
17 tools utilized by Etilize.

18
19 D. Computational Linguistics

20 The final question is whether Etilize literally, or under the doctrine of equivalents, infringes
21 the computational linguistics limitation of Claims 1 and 15 of the ‘933 Patent. The Claim limitation
22 states that the “crawler utilizes computational linguistics to gather said product phrase information
23 which includes a phrase and at least one characteristic of said phrase.” ‘933 Patent at 18:63–65.

24 A “phrase” is a “string of characters, such as an alpha-numeric character string or strings
25 present in a source.” Docket No. 66, Exh. A at 1. Further, “phrase characteristics” is defined as:

26 Some attribute of a phrase that can be used to distinguish it from other phrases, such
27 as the frequency, location, font size, font style, font case, font effects, and font color
28 as well as the frequency of collocation (phrases immediately next to each other) and
co-occurrence of phrases (phrases within a predetermined number of words of each

1 other) includes the co-occurrence of phrases, which is when a particular word is
2 within a number of words of another (e.g., ‘weight’ and ‘lbs.’).

3 Id. Based on the above, if the string of characters “weight: 5.8 lbs” appears in a source and that
4 string is gathered, then “product phrase information” could have been gathered. Specifically, the co-
5 occurrence of the phrase “weight” and the phrase “lbs” within a few words of each other
6 distinguishes the phrase “weight” from other phrases. Thus, the “at least one characteristic of said
7 phrase” limitation is met by the string through co-occurrence. Further, there can be no argument that
8 the string of characters constituting the word “weight” is a phrase. Consequently, since both a
9 “phrase” and “at least one characteristic of said phrase” have been gathered, “product phrase
10 information” could have been gathered. There is also no argument that Xtract gathers exactly this
11 type of information.

12 The only remaining question is whether “computational linguistics” was utilized to gather the
13 product phrase information. “Computational linguistics” is “[a] cross disciplinary field of modeling
14 of language utilizing computational analysis to process language data.” Id. CNET’s expert testifies
15 that the definition of “computational linguistics” encompasses the use of regular expressions which
16 he contends “use computational analysis to process text language data and are a way to describe text
17 through pattern matching.” Gray Dec., Exh. A at 36.

18 CNET contends Xtract practices this limitation by its use of regular expressions. A regular
19 expression is a string of characters that is used to describe or match a set of strings according to
20 certain rules used to construct sentences in natural languages. The specific rules of construction
21 vary depending on the task, but regular expressions can search, manipulate and process text-based
22 patterns. Language data, e.g., “weight: 5.8 lbs” can be matched according to a regular expression
23 that looks for the following pattern: the string “weight: #.#” followed immediately by either the
24 string “lbs” or “oz” and where # stands for any string of numerical characters only. Computational
25 analysis would be required to parse a phrase to determine if it matches the language data sought.
26 Thus, this modeling of parsed phrases could meet the definition of computational linguistics.
27 Accordingly, CNET’s expert is correct in concluding that regular expressions could be encompassed
28 within the construction of computational linguistics. Alternatively, even if Etilize does not use

1 regular expressions, the court finds that use of pattern matching to find relevant phrases on a web
2 page could meet the definition of computational linguistics. This is because pattern matching
3 utilizes computational analysis (comparing letters in a string) to process language data (phrases in
4 the web page).

5 Etilize incorrectly argues that because it utilizes a manual process, and not a crawler, it
6 cannot carry out the computational linguistics required by Claim 1. Etilize thus converges two
7 separate inquiries into one because if the court finds use of a crawler, the computational linguistics
8 issue remains ripe. Because of this conflation, Etilize correctly argues only that aQuire does not
9 carry out computational linguistics and thus does not infringe the '933 Patent. It, however, ignores
10 the parsing done by Xtract.³

11 Etilize acknowledges that Xtract pulls or gathers attribute/value pairs from the collected web
12 pages using predefined expression patterns.⁴ Hameed Dec. ¶ 12. Based on this admission, the court
13 finds it unnecessary to rely upon the Christensen declaration, which points out the deficiencies in
14 CNET's expert report. Indeed, even though Etilize asserts that much of CNET's expert declaration
15 is deficient because it analyzes source code that is not used by Xtract, nowhere does Etilize state that
16 Xtract does not make use of predefined expressions. Further, instead of refuting the argument that
17 use of predefined expressions constitutes use of computational linguistics, Etilize argues that
18 because the use of Xtract is not fully automated, it cannot utilize computational linguistics. When
19 making this argument Etilize states that since "researchers have to specify for every product attribute
20 where to extract the information from the web page, what information to extract, and where to put it
21 in the template," *id.*, Xtract does not utilize computational linguistics. This argument fails for two
22 reasons. First, the court has already rejected Etilize's proposition that two additional
23 limitations—"automatically" and "without human intervention"—should be inserted into the term
24 "crawler." Docket No. 82 at 10. Thus, simply because some, or even a majority, of the operations
25 Etilize performs to gather product information are performed manually does not necessitate that
26 Xtract cannot utilize computational linguistics. Second, although researchers have to specify where
27 to extract product information from a web page, what information to extract and where to put it in
28 the template, the actual obtaining or gathering of the attribute/value pairs, e.g., "attribute: weight;

1 value: ## lbs,” is nevertheless performed by Xtract through the use of predefined expression
2 patterns.

3 The expression patterns used by Xtract can be used to gather product phrase information
4 because they gather a phrase and at least one characteristic of said phrase as required by Claim 1 of
5 the ‘933 Patent. For example, where a string of characters with the values of “price,” “hard drive
6 capacity,” and “weight” are found on a website, regular expressions may be used to produce the
7 following results: “price: \$1299.99;” “hard drive: 40GB;” and “weight: 5.4 lbs.”

8 In sum, it is the use of expression patterns that may constitute the use of computational
9 linguistics to gather product phrase information.

10 Etilize argues that Xtract does not gather any formatting or other characteristics of the
11 extracted attributes; instead, it simply allows a research operator to “pull” attribute/value pairs from
12 isolated web pages for entry into the template. These two statements by Etilize are inconsistent
13 because “pulling” attribute/value pairs is synonymous with “gathering” those pairs. Since pulling
14 name/value pairs of data through pattern matching can be the utilization of computational linguistics
15 to gather product phrase information, the court concludes that CNET has met its burden and this
16 limitation may be practiced by Etilize.⁵

17 In sum, since the software tools employed by Etilize, aQuire and Xtract, could, in
18 combination, be found to practice every limitation set forth in Claims 1 and 15 of the ‘933 Patent,
19 summary judgment on non-infringement is DENIED.

20
21 **III. Infringement of the ‘426 Patent**

22 Etilize argues its software tools do not infringe the claims of the ‘426 Patent because they do
23 not practice the following limitations: 1) grouping; 2) electronic comparing; 3) electronic parsing;
24 4) electronic generating; 5) use of a crawler; 6) crawler generated from a server; and 7) visiting a
25 plurality of sources. Limitations five through seven are discussed above and the remaining
26 limitations, to the extent necessary, are discussed here. However, the court first discusses Etilize’s
27 contention that it does not perform the claim limitations in the order specified in the claims.

28

1 A. Order of Performance

2 The court accepts Etilize’s argument that the claims at issue here cannot be infringed unless
3 the accused product practices each limitation in the same order as stated in the claim. CNET cites to
4 Loral Fairchild Corp. v. Sony Corp., 181 F.3d 1313, 1322 (Fed. Cir. 1999), to argue that order is
5 irrelevant if all the limitations are practiced by the accused device. In Loral, the court concluded
6 that “although not every process claim is limited to the performance of its steps in the order written,
7 the language of the claim, the specification and the prosecution history [may] support a limiting
8 construction.” Id. at 1321. Additionally, where the “sequential nature of the claim steps is apparent
9 from the plain meaning of the claim language and nothing in the written description [of the patent]
10 suggests otherwise” the steps of the claim are to be performed in sequential order. Mantech Envtl.
11 Corp v. Hudson Envtl. Servs., Inc., 152 F.3d 1368, 1375–76 (Fed. Cir. 1998). Here, the plain
12 meaning of the claim language supports a finding that the steps within Claims 1, 39 and 52 are to be
13 performed sequentially.

14 Claim 1 of the ‘426 Patent recites:

15 A method of creating a product catalog stored on computer readable media by
16 aggregating product information from a plurality of product information sources
17 having disparate formats for product information and storing the information in a
18 taxonomy, said method comprising:

19 processing plural product information records from the product information
20 sources into one or more groups based on which product information
21 records are likely to correspond to the same product;

22 correlating a unique product ID corresponding to the product associated
23 with each of said groups to identify the product;

24 electronically comparing each identified product to categories of a
25 taxonomy to determine a category for the identified products in the
26 taxonomy; and

27 electronically parsing the product information records corresponding to each
28 group to electronically determine attributes for each categorized product
based on the product information records;

 electronically generating product specifications based on the determined
attributes; and

 storing the product specification in the corresponding determined categories
of the taxonomy.

‘426 Patent at 36:22–45.

1 Claim 1 must be performed sequentially. The initial step, the processing step, creates the
2 groups. The second step correlates a unique product ID to each group. No correlation of a product
3 ID to the groups can occur until the groups are created. The third step compares the unique product
4 to categories of a taxonomy. However, no categorization can occur until the product is identifiable
5 by a product ID. The fourth step determines the categorized product's attributes. This obviously
6 requires that the product be categorized. The fifth step generates product specifications based on the
7 determined attributes. However, no specifications could be generated until the attributes are
8 determined in the fourth step. Finally, the product specification is stored in the previously
9 determined category of the taxonomy. No product specification could be stored until it is created in
10 the previous step. Thus, the steps of this method claim must be performed in order. Because Claim
11 39 is substantially similar to Claim 1, the same conclusion follows for Claim 39.⁶

12 Etilize's process categorizes the product in question *before* any of the steps in the claims are
13 performed. Specifically, after Etilize determines that information about a particular product must be
14 acquired, it first classifies the product to a category. Hamed Dec. ¶ 8. This classification can be
15 manual or automatic. *Id.* Since each product category has a template associated with it, upon
16 association of a product with a category, a product template with relevant product attributes is also
17 associated with the product. *Id.* ¶ 9. For instance, the template for the category of notebook
18 computer would include relevant product attributes such as processor technology, RAM memory
19 space, hard drive capacity and display size. If no template exists for the category, one is created
20 manually. *Id.* The template could then be populated using aQuire and Xtract. Claims 1 and 39 of
21 the '426 Patent requires that the grouping step occur before the product is categorized; however, the
22 Etilize process categorizes before grouping.

23 CNET's expert simply asserts that ConQuire, another software tool utilized by Etilize,
24 compares the product to categories of a taxonomy to determine the appropriate category.⁷ That may
25 be true, however, the same does not demonstrate that the grouping occurs before the categorization.
26 Indeed, the use of ConQuire, to the extent it performs categorization, is perfectly consistent with
27 Hameed's declaration, which states that the categorization is sometimes performed automatically.
28 *Id.* ¶ 8. With respect to the order in which the steps are performed, CNET's expert simply states that

1 “Etilize’s asserted different order, or combination of multiple steps into a single step is an
2 insubstantial change from the process elements of Claim 1, and any such change performs
3 substantially the same function in substantially the same way to achieve substantially the same
4 result.” Gray Dec., Exh. A at 47. This conclusory statement cannot carry the day. Consequently,
5 the court finds that Etilize’s process categorizes the product before performing any of the steps of
6 Claims 1 and 39 of the ‘426 Patent. Thus, Etilize does not infringe Claims 1 and 39, as well as all
7 claims dependent on these independent claims of the ‘426 Patent.

8
9 B. Grouping

10 The “grouping” limitation is present in the first limitation of Claims 1 and 39 of the ‘426
11 Patent as well as in the third limitation of Claim 52 of the ‘426 Patent. Etilize argues that it does not
12 sort or otherwise process a collection of product information records into groups based on their
13 similarities or differences.

14 The use of aQuire “necessitates that there [be] a vendor with a large number of similar
15 products being acquired and a URL pattern [that] can be repeated over and over again [by] just
16 replacing the product ID or SKU.” Hameed Dec., Exh. D at 5. Hameed states that if, for example,
17 the “product category is disk drive systems available from IBM and the IBM site generally has three
18 pages that are associated with each product” aQuire would store copies of all three pages associated
19 with the product on an Etilize server. Hameed Dec. ¶ 11. For instance, to fetch pages related to
20 IBM’s x500 disk drive, aQuire would copy: “ibm.com/disk_drive/x500/index.html,”
21 “ibm.com/disk_drive/x500/features.html” and “ibm.com/disk_drive/x500/specifications.html.”

22 According to CNET, aQuire practices this limitation because it downloads multiple web
23 pages associated with a singular product and these downloaded web pages constitute a “group.”
24 CNET’s expert states that “[p]rior to extraction from the data from the web page [by Xtract] the
25 process method assigns the [downloaded] page to a group based on the *site data related to the link*
26 that was crawled to obtain the page.” Gray Dec., Exh. A at 44 (emphasis added). No explanation is
27 given as to the “site data related to the link.” Consequently, the court is left to conjecture. The court
28 finds that “site data related to the link” simply means that the uniform resource locator (“URL”) of

1 the website is used to identify the downloaded web page. Simply stated, Etilize’s process likely
2 stores the website www.ibm.com/disk_drive/x500/index.html on its servers with the filename
3 “ibm.com_disk_drive_x500_index.html.”⁸ Other downloaded web pages associated with the x500
4 disk drive likely have similar names. It is this similarity of identification that CNET presumably
5 claims constitutes the group.

6 This so-called grouping, however, is fundamentally different from the grouping envisioned
7 by the claim limitation. The limitation speaks to “processing plural product information records
8 from the product information sources into one or more groups based on which product information
9 records are likely to correspond to the same product.” Here, it is uncontested that each web page,
10 which CNET implicitly contends is a “product information record,” refers to a particular known
11 product. No determination as to which web page belongs to which product needs to be made. This
12 information is known well before the web page is even downloaded. Consequently, the notion that
13 web pages need to be placed into a group such that all group members likely all correspond to the
14 same product is nonsensical.

15 CNET is unsuccessful even if the court accepts that Etilize’s process requires that web pages
16 be placed into groups based on which product information records, i.e., web pages, are likely to
17 correspond to the same product. An argument can be made that all downloaded web pages with
18 “x500” in their filename belong to a group that represents a unique product, likely IBM’s x500 disk
19 drive. This grouping based on the “site data related to the link,” however, is not performed by
20 Etilize to identify the product or its attributes, as required by the claim limitations. The so-called
21 grouping is based on URL’s associated with an identified product that already has predefined
22 attributes in Etilize’s template, and therefore, no reasonable jury could find that it is performed with
23 the purpose of either identifying the product, ‘426 Patent, Claims 1, 39, or identifying the attributes
24 of the product, ‘426 Patent, Claim 52.

25 CNET argues that grouping should extend not only to the grouping of product information
26 records based on which records are likely to correspond to the same product, but also to the grouping
27 and prioritizing of Etilize’s customer requests. Gray Dec., Exh A at 45. By broadly interpreting the
28 grouping limitation, not only does CNET seek to now expand upon the construction that the parties

1 agreed to but also to recapture what it disavowed during the prosecution of the ‘426 Patent. In order
2 to differentiate the ‘426 Patent from the prior-art reference of Blutinger, CNET specified that “the
3 recited initial processing of the product information records into one or more groups does not fall
4 under any taxonomy structure. Instead, this grouping refers to the fact that the product information
5 records are analyzed for similarities and differences and associated together based upon which
6 product information records are likely to correspond to the same product.” Khaliq Dec., Exh. G at
7 27. Based on this prior disavowal and the parties’ agreement to the contrary, the court rejects
8 CNET’s broad interpretation and restricts the limitation to associating together product information
9 records that are likely to correspond to the same product.⁹

10 In sum, Etilize’s products do not practice this limitation and summary judgment on non-
11 infringement of every independent, and consequently dependent, claim of the ‘426 Patent still in
12 issue in this action is GRANTED.¹⁰

13
14 C. Burden Shifting

15 CNET argues that summary judgment is premature because the court has not yet addressed
16 the applicability of 35 U.S.C. section 295. Section 295 states:

17 In actions alleging infringement of a process patent based on the importation, sale,
18 offer for sale, or use of a product which is made from a process patented in the United
19 States, if the court finds–

- 19 (1) that a substantial likelihood exists that the product was made by the
20 patented process, and
21 (2) that the plaintiff has made a reasonable effort to determine the process
22 actually used in the production of the product and was *unable so to determine*,
23 the product shall be presumed to have been so made, and the burden of establishing
24 that the product was not made by the patented process shall be on the party asserting
25 that it was not so made.

26 35 U.S.C. § 295 (emphasis added). Here, CNET has been able to determine the process actually
27 used in the production of the allegedly infringing product. Indeed, CNET has all of Etilize’s source
28 code. Thus, because the second factual requirement is not met, the burden shifting of section 295
does not apply.

1 IV. Motions to Strike

2 CNET objects to the declarations submitted by Etilize’s Chief Executive Officer (“CEO”)
3 and an individual in its employ. Specifically, CNET argues that: 1) Etilize’s CEO, Azhar Hameed,
4 is not competent to testify because he has no personal knowledge of the software at issue; and
5 2) Etilize’s employee, Benjamin Christensen, is not competent to testify because he is introducing
6 new arguments for the first time in Etilize’s reply and because of a lack of personal knowledge.

7
8 A. Azhar Hameed

9 CNET makes three specific arguments. First, Hameed testified that he has neither seen nor
10 used the Xtract system; second, Hameed has no familiarity with Webspinx; and third, Hameed did
11 not prepare the entirety of his declaration on his own. In the alternative, CNET seeks to strike
12 paragraphs 1, 5 and 7–14 of Hameed’s declaration. Each of these arguments is unconvincing.

13 First, CNET is correct to state that Hameed simply states, in a conclusory manner, that his
14 declaration is based on personal knowledge. He does not specify his basis for the same. However,
15 Hameed is Etilize’s CEO. In his capacity as CEO, he must have knowledge as to how his company
16 conducts business. Specifically, the fact that Hameed has never used the software in question is of
17 no occasion. Hameed does not purport to explain the details of the software or testify as an expert
18 witness. He simply seeks to explain the general process employed by the software. He need not run
19 the software himself in order to understand the broad strokes; he may rely upon subordinates to
20 explain the same to him. Indeed, as CEO, it is proper for the court to infer that Hameed has general
21 personal knowledge of the process used to obtain data for Etilize’s product. See Barthelemy v. Air
22 Lines Pilots Ass’n, 897 F.2d 999, 1018 (9th Cir. 1990) (“personal knowledge and competence to
23 testify are reasonably inferred from [the declarant’s] position[] and the nature of their participation
24 in the matters to which they swore”).

25 Second, the court does not consider Hameed’s testimony with respect to Webspinx and
26 consequently, objections to the evidence on that issue are moot.

27 Third, the fact that Hameed may have copied, verbatim, language from opinion counsel’s
28 opinion is also unpersuasive. Hameed swore to the functionality of the software under penalty of

1 perjury and if opinion counsel’s language was the clearest formulation of the same, then the court
2 will not penalize Hameed for recycling the language.

3 CNET also argues that Hameed’s declaration contains conclusions that are unsupported by
4 underlying factual information. To the extent that is true, it pertains to the evidentiary weight of the
5 declaration, not the admissibility thereof.

6 In sum, CNET’s motion to strike Hameed’s declaration is DENIED. For the same reasons,
7 CNET’s alternative request to strike particular paragraphs of Hameed’s declaration is also DENIED.

8
9 B. Benjamin Christensen

10 The court does not rely upon the Christensen’s declaration in reaching its decision.
11 Consequently, all of CNET’s objections to Christensen’s declaration are DENIED as moot.

12
13 V. Motion to Bifurcate

14 Etilize seeks to bifurcate this action into two phases. Specifically, it seeks resolution of
15 whether Etilize infringes CNET’s patents and whether those patents are valid and enforceable before
16 resolution of whether CNET has engaged in unfair business practices. The motion is denied as
17 untimely for the following reasons.

18 Federal Rule of Civil Procedure 42(b) states: “For convenience, to avoid prejudice, or to
19 expedite and economize, the court may order a separate trial of one or more separate issues, claims,
20 crossclaims, counterclaims, or third-party claims.” Etilize has raised some compelling reasons to
21 bifurcate the trial in this action: 1) bifurcation will reduce the risk of jury confusion; 2) the issues are
22 separable; and 3) bifurcation will lead to judicial economy. However, the trial date for this action is
23 still at least six months away. See Docket No. 86. Indeed, expert discovery is ongoing and there are
24 recently filed dispositive motions that the court has yet to be heard. Thus, the issues that will
25 actually be tried may be vastly different from the issues currently present in the action.

26 Consequently, the court shall not order the trial bifurcated at this time.

27 The court notes that even if it ordered the trial bifurcated, the same does not necessitate that
28 discovery on the bifurcated issues be stayed. Indeed, the motion to bifurcate seems designed by

1 Etilize to simply delay discovery on counterclaims brought by Etilize. Etilize states it does not wish
2 to incur the expense necessary to prepare the expert report for its unfair business practices cause of
3 action. However, Etilize brought the counterclaim and cannot now be heard to argue that bifurcation
4 will save it the expense of preparing expert reports. Furthermore, none of the arguments Etilize
5 advances in support of bifurcation support a stay of discovery regarding the issues to be bifurcated.
6 For this additional reason, the motion to bifurcate is DENIED. While bifurcation of discovery is not
7 justified, the court may at a later date consider bifurcation for trial.

8
9 VI. Etilize's Supplemental Brief

10 The court does not rely upon the arguments regarding the '426 Patent in Etilize's
11 supplemental brief since the same is unnecessary. Further, the court does not rely upon Etilize's
12 arguments regarding the "determining" step of the '933 Patent since the same was not put into issue
13 by Etilize in its opening brief.¹¹ Further, considering Etilize's submission would unfairly prejudice
14 CNET by allowing Etilize to place new information and argument before this court that CNET has
15 not had an opportunity to refute. To the extent that the supplemental brief relies upon new
16 information previously unavailable to Etilize, Etilize may not be heard to complain. Etilize filed the
17 instant motion on non-infringement before expert discovery was complete. Consequently, to the
18 extent Etilize did not have the relevant information at the time it filed its motion, the ignorance was
19 self-inflicted.

20
21 CONCLUSION

22 For the foregoing reasons, defendant's motion on noninfringement is GRANTED in part and
23 DENIED in part; plaintiff's motions to strike are DENIED; defendant's motion to bifurcate is
24 DENIED; and defendant's motion to file a supplemental brief is DENIED.

25 IT IS SO ORDERED.

26 Dated: August 29, 2008

27 
28 MARILYN HALL PATEL
United States District Court Judge
Northern District of California

ENDNOTES

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

1. Defendant’s request to file an oversized reply brief is GRANTED.
2. The difference between the two correlating steps is insubstantial. Compare “correlating a unique product ID corresponding to an identified product for each of said groups” with “correlating a unique product ID corresponding to the product associated with each of said groups to identify the product.” ‘426 Patent at 40:8–9; 36:31–33.
3. The fact that Xtract, which uses html identifiers, is run on a downloaded copy of a web page as opposed to the same web page on the internet is of no significance. Running Xtract on either results in the same functions being carried out in the same way to achieve the same result.
4. The fact that the extraction takes place from tables or predefined coordinates on a web page is irrelevant since pattern matching is nevertheless used to perform the extraction.
5. The analysis for finding infringement as a matter of law requires that all claim limitations be practiced by the accused product. CNET has not moved for the same here.
6. Claim 52 also requires sequential performance. The first step generates the crawler, the second uses the crawler to gather information, the third groups this information, the fourth parses the grouped information to determine attributes and the fifth creates a catalog based on the determined attributes. The steps must be performed in order because each subsequent step depends upon the results generated by its predecessor.
7. The fact that ConQuire is not itself accused of infringement is irrelevant because SpeX, the final product sold by Etilize and which utilizes ConQuire, is the product accused of infringement.
8. Depending on the operating system used, the software also likely changes the URL slightly to account for special characters that may not be used as part of a filename.
9. CNET also argues that placing web pages to be downloaded via aQuire in a queue is “grouping.” This does not constitute a “group” in any sense of the word. Indeed, it is impossible to determine the group members. Furthermore, to the extent this queuing could constitute “grouping” as envisioned by the claims of the ‘426 Patent, it is not performed for the purpose of identifying the product or its attributes.
10. Based on this holding, the court does not reach the parties’ “electronically” arguments.
11. The same rationale applies to the argument regarding the “determining” step in Etilize’s reply brief.