

IN THE UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF DELAWARE

XEROX CORPORATION,	)	
	)	
Plaintiff-Counterclaim Defendant,	)	
	)	
v.	)	C.A. No. 10-136-LPS
	)	
GOOGLE INC., YAHOO! INC., RIGHT MEDIA	)	
INC., RIGHT MEDIA LLC, YOUTUBE, INC. AND	)	
YOUTUBE, LLC,	)	
	)	
Defendants-Counterclaim Plaintiffs.	)	

**PLAINTIFF XEROX CORPORATION'S  
RESPONSIVE CLAIM CONSTRUCTION BRIEF**

ASHBY & GEDDES  
Lawrence C. Ashby (I.D. #468)  
John G. Day (I.D. #2403)  
Lauren E. Maguire (I.D. #4261)  
500 Delaware Avenue, 8th Floor  
P.O. Box 1150  
Wilmington, DE 19899  
(302) 654-1888

*Attorneys for Plaintiff Xerox Corporation*

*Of Counsel:*

Richard J. Stark  
Andrei Harasymiak  
CRAVATH, SWAINE & MOORE LLP  
Worldwide Plaza  
825 Eighth Avenue  
New York, NY 10019-7475  
(212) 474-1000

Dated: April 26, 2011

**TABLE OF CONTENTS**

	<b>Page</b>
TABLE OF AUTHORITIES .....	ii
I. INTRODUCTION .....	1
II. LEGAL BACKGROUND .....	2
III. DEFENDANTS IMPROPERLY ATTEMPT TO ENGRAFT ONTO THE CLAIMS A SET OF LIMITATIONS AS TO HOW SEARCHES ARE IMPLEMENTED .....	3
A. The Single-Classification-Label Fallacy.....	3
B. The Request-for-Search-Results Fallacy .....	5
C. The Fallacy of Substituting “Confine” for “Restrict” .....	8
IV. THE FORMAT-OF-THE-LABEL FALLACY .....	12
V. DEFENDANTS FAIL TO SHOW THAT THE TERM “SELECTED DOCUMENT CONTENT” IS INDEFINITE.....	16
VI. DEFENDANTS’ REMAINING CLAIM CONSTRUCTION ARGUMENTS ARE WITHOUT MERIT .....	18
A. “Characteristic Vocabulary” Describes a Class in the Organized Classification of Document Content .....	18
B. Nothing in Claims 1 and 18 or the Specification Requires the “Automatically Identifying” and “Automatically Categorizing” Steps To Be Fully Completed before the “Automatically Formulating the Query” Step Is Commenced .....	18
C. Nothing in the Claims or the Specification Requires the Steps of Claims 1 and 18 To Be Fully Completed before the Steps of Claims 2 and 19 Are Commenced .....	20
VII. CONCLUSION.....	20

**TABLE OF AUTHORITIES**

	<b>Page(s)</b>
<b>Cases</b>	
<i>Altiris, Inc. v. Symantec Corp.</i> , 318 F.3d 1363 (Fed. Cir. 2003) .....	19
<i>Baldwin Graphic Sys., Inc. v. Siebert, Inc.</i> , 512 F.3d 1338 (Fed. Cir. 2008) .....	3, 5, 18
<i>Bell Comms. Research, Inc. v. Fore Sys., Inc.</i> , 62 Fed. Appx. 951, 2003 WL 1720080 (Fed. Cir. 2003) .....	19
<i>Broadcom Corp. v. Qualcomm Inc.</i> , 543 F.3d 683 (Fed. Cir. 2008) .....	6, 7
<i>Chimie v. PPG Indus., Inc.</i> , 402 F.3d 1371 (Fed. Cir. 2005) .....	11
<i>Energizer Holdings, Inc. v. Int’l Trade Comm’n</i> , 435 F.3d 1366 (Fed. Cir. 2006) .....	16
<i>Halliburton Energy Servs., Inc. v. M-I LLC</i> , 514 F.3d 1244 (Fed. Cir. 2008) .....	16, 17
<i>Howmedica Osteonics Corp. v. Wright Med. Tech., Inc.</i> , 540 F.3d 1337 (Fed. Cir. 2008) .....	2
<i>i4i Ltd. P’ship v. Microsoft Corp.</i> , 598 F.3d 831 (Fed. Cir. 2010) .....	11
<i>Invitrogen Corp. v. Biocrest Mfg., L.P.</i> , 327 F.3d 1364 (Fed. Cir. 2003) .....	3
<i>Liebel-Flarsheim Co. v. Medrad, Inc.</i> , 358 F.3d 898 (Fed. Cir. 2004) .....	2, 9
<i>Osram GmbH v. Int’l Trade Comm’n</i> , 505 F.3d 1351 (Fed. Cir. 2007) .....	20
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) .....	passim
<i>Praxair, Inc. v. ATMI, Inc.</i> , 543 F.3d 1306 (Fed. Cir. 2008) .....	16
<i>SRI Int’l v. Matsushita Elec. Corp. of Am.</i> , 775 F.2d 1107 (Fed. Cir. 1985) .....	2
<i>Ventana Med. Sys., Inc. v. Biogenex Labs., Inc.</i> , 473 F.3d 1173 (Fed. Cir. 2006) .....	6, 7
<i>Wyeth, LLC v. Intervet, Inc.</i> , C.A. No. 09-161-LPS, 2011 WL 1043575 (D. Del. Mar. 22, 2011) .....	2, 4, 9, 14, 20
<i>Xerox Corp. v. Google, Inc.</i> , Civ. A. No. 10-136 (D. Del. Mar. 30, 2011) .....	10
<b>Statutes</b>	
35 U.S.C. § 112 .....	6, 14

## I. INTRODUCTION

Defendants' claim construction arguments proceed from a number of false and legally incorrect premises and should be rejected. The principal dispute here centers on whether Defendants can import into the claims the requirement that a query can only be a request for search results confined to precisely one category of information. (*See* Defs.' Opening Br. ("Def. Br."), D.I. 141, at Argument §§ I, II and IV.) Defendants construct their argument on three major fallacies: (1) that the singular form "a classification label" limits the claims to using exactly one classification label; (2) that claims directed solely at query formulation nevertheless limit the potential operation of the formulated query; and (3) that a classification label must "confine" the scope of the formulated query to a single category of information in the precise manner shown in a specification example. As to the first proposition, the law is clear that in patent claims "a" means "one or more" absent exceptional circumstances not present here. In fact, the specification expressly states that categorization can result in more than one classification label, and that more than one label may be used in a query. Defendants' second proposition is flatly inconsistent with the claim language. Indeed, Defendants do not even attempt to justify their construction by citing the independent claims in suit, but instead resort to irrelevant dependent claims. Finally, Defendants' third proposition ignores the basic canon of claim construction that a specification example cannot be used to limit the scope of a claim absent clear evidence of the patentee's restrictive intent. Here, the specification never teaches that a formulated query can only invoke a search in a single "folder" or category of information. To the contrary, the specification expressly teaches that any search need only be "focused" on particular categories of information. Defendants' proposed constructions conflict with the intrinsic evidence, and their attempts to mask these contradictions by resorting to extrinsic

evidence are untenable. For all these reasons, Defendants' primary claim constructions are incorrect and should be rejected.

Defendants' remaining positions fare no better. Their insistence that a classification label must be a word or phrase again seeks to treat an example in the specification as though it were a claim limitation. Moreover, it ignores the basic fact that computers are digital machines in which information is always translated into numeric data. Defendants' indefiniteness and order-of-steps arguments clearly fail under controlling case law, while their construction of "characteristic vocabulary" conflicts with both the claim language and the specification.

## II. LEGAL BACKGROUND

As this Court has recognized, "[i]t bears emphasis that '[e]ven when the specification describes only a single embodiment, the claims of the patent will not be read restrictively unless the patentee has demonstrated a clear intention to limit the claim scope using words or expressions of manifest exclusion or restriction'." *Wyeth, LLC v. Intervet, Inc.*, C.A. No. 09-161-LPS, 2011 WL 1043575, at \*3 (D. Del. Mar. 22, 2011) (quoting *Liebel-Flarsheim Co. v. Medrad, Inc.*, 358 F.3d 898, 906 (Fed. Cir. 2004)). Of course, the intrinsic evidence, mainly the claim language itself and the specification, dominate claim construction analysis. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1314-15 (Fed. Cir. 2005). The prosecution history can provide assistance in certain situations. *Id.* at 1317. Some extrinsic evidence, such as dictionaries, can be of limited use. *Id.* at 1317-19. Other extrinsic materials, such as inventor testimony regarding his intent, have no relevance.<sup>1</sup> *Howmedica Osteonics Corp. v. Wright Med. Tech., Inc.*, 540 F.3d

---

<sup>1</sup> In addition, the Federal Circuit has held that "[i]t is only *after* the claims have been construed without reference to the accused device that the claims, as so construed, are applied to the accused device to determine infringement." *SRI Int'l v. Matsushita Elec. Corp. of Am.*, 775 F.2d 1107, 1118 (Fed. Cir. 1985) (emphasis in original). Defendants' conclusory (and inaccurate) attempts to characterize the operation of their products and prematurely argue infringement issues are thus improper and irrelevant.

1337, 1347 (Fed. Cir. 2008) (“inventor testimony as to the inventor’s subjective intent is irrelevant to the issue of claim construction”).

### **III. DEFENDANTS IMPROPERLY ATTEMPT TO ENGRAFT ONTO THE CLAIMS A SET OF LIMITATIONS AS TO HOW SEARCHES ARE IMPLEMENTED**

Defendants principally contend that the invention of Claims 1 and 18 must yield a formulated query that (1) must contain one and only classification label; (2) must be a “request for search results”; and (3) must “confine” any search to a single category of information. This central contention rests on three critical fallacies. We address each below.

#### **A. The Single-Classification-Label Fallacy**

Defendants contend that the phrase “a classification label” in the “automatically categorizing” step of Claims 1 and 18 means **a single** classification label. (Def. Br. at 11-12.) As Defendants recognize, this is “[t]he crux of the parties’ dispute” (Def. Br. at 11), and it is the cornerstone of their argument that the claims require a search in a single category of information.

Binding Federal Circuit precedent rejects Defendants’ interpretation. The Federal Circuit has stated in no uncertain terms:

That “a” or “an” can mean “one or more” is best described as a rule, rather than merely as a presumption or even a convention. The exceptions to this rule are extremely limited: a patentee must evince a clear intent to limit “a” or “an” to “one”. *Baldwin Graphic Sys., Inc. v. Siebert, Inc.*, 512 F.3d 1338, 1342-43 (Fed. Cir. 2008) (internal quotations omitted).

Thus, absent strong evidence that the patentee intended “a” to mean “one”, “a” means “one or more”.<sup>2</sup>

---

<sup>2</sup> Although interpreting “a” to mean “one or more” may seem counterintuitive, this result follows logically from the rule that an open-ended patent claim (a claim with the transitional term “comprising”) does not “exclude additional, unrecited elements or method steps.” MPEP § 2111.03 (Ex. 1). *See also Invitrogen Corp. v. Biocrest Mfg., L.P.*, 327 F.3d 1364, 1368 (Fed. Cir. 2003). Thus, if an open-ended claim requires “a widget,” it is satisfied as long as there is one widget; the presence or use of additional widgets is not excluded by the claim. Or, in the context of the “automatically categorizing” step of open-ended Claims 1 and 18 of the ’979 Patent, if categorization results in the assignment of at least “a” classification label, the claim

There is no such evidence here. Rather, the evidence goes the other way. The '979 specification clearly states:

Document *classification labels* define the *set of categories* 3620 output by the categorizer 3610. These *classification labels* in one embodiment are appended to the query 3812 by query generator 3810 to restrict the scope of the query (i.e., the entity 3808 and the context vector 3822) to folders corresponding to *classification labels* in a document collection of an information retrieval system. (Xerox Opening Br. (“X. Br.”), D.I. 142, at 10 (citing 979/49:31-37) (emphasis added).)

The specification further states that “[i]n generating the *set of categories* 3620, the categorizer 3610 classifies input document to generate *classification labels* for the document content 3612”. (979/49:18-20 (emphasis added).) Indeed, the specification expressly and repeatedly teaches that document content can be categorized into multiple categories. (See X. Br. at 10-11; see also 979/48:40-48; 979/41:53-58; 979/43:35-40.)

Defendants’ reliance on an out-of-context sentence from a prior decision of this Court, *Wyeth*, does not avail them. *Wyeth* construed the claim term “porcine circovirus type B” (“PCVB”). 2011 WL 1043575, at \*4. The question was whether PCVB should be construed as a single virus or a group of related viruses. *Id.* The Court undertook an extensive analysis of the use of various terms in the specification to determine that PCVB referred to one particular virus. *Id.* at \*4-\*9. As one part of that analysis, the Court noted that the specification referred to “the genome of PWD circovirus type B”, which apparently referred to a singular genome of a singular virus. *Id.* at \*6. Notably, the Court was not construing the words “a” or “the” in the patent claim, but rather pointing out that the specification’s description of the invention linked the term PCVB to a single virus. The holding in *Wyeth* is therefore inapposite here.

Defendants cite the use of the definite article “the” with “classification label” later in Claims 1 and 18 to support their interpretation of “a” as exactly “one”. The Federal Circuit has

---

language is satisfied. Since the assignment of additional classification labels is neither required nor forbidden by the claims, the claim effectively covers “one or more” classification labels.

definitively rejected that argument. “A” means “one or more”, and “[t]he subsequent use of definite articles ‘the’ or ‘said’ in a claim to refer back to the same claim term does not change the general plural rule, but simply reinvokes that non-singular meaning.” *Baldwin*, 512 F.3d at 1342. Defendants also argue that the use of “the category” in the phrase “the category of information in the information retrieval system identified by the assigned classification label” proves that there is only one assigned classification label. But this argument ignores the Federal Circuit law and specification teachings cited above. Although “the assigned classification label” is grammatically singular, it is presumptively plural as a matter of law, an interpretation repeatedly confirmed by the specification of the ’979 Patent. In using one singular noun phrase (“the category of information”) with another (“the assigned classification label”), the claim language simply follows the conventions of English usage. This is hardly the “clear intent to limit ‘a’ or ‘an’ to ‘one’” that the controlling law requires (*see Baldwin*, 512 F.3d at 1342), especially in view of the specification’s repeated teaching that multiple categories and labels may be used.

Both the specification and binding case law confirm that “a classification label” means “one or more classification labels.” Defendants’ construction must therefore be rejected.

### **B. The Request-for-Search-Results Fallacy**

Defendants also attempt to narrow the scope of the asserted claims by ignoring their subject matter. Claims 1 and 18 address “generating a query”. Both claims specify data that must appear in the query (data corresponding to entities and classification labels). Both claims conclude with the step of “automatically formulating the query” that contains that data. But any interaction of the formulated query with an information retrieval system would come later. Claims 1 and 18 do not go that far. They cover only the formulation of a query, full stop. Accordingly, Defendants’ attempts to require that a query comprise only a “request for search results” have no foundation in the claim language.

Defendants assert that “as the claims make clear, a query is used to ‘search.’ The claims also talk about the query being ‘performed’ and used to obtain ‘results.’” (Def. Br. at 14.) To support this contention, Defendants point not to Claims 1 and 18, but rather to two *dependent* claims (Claims 11 and 17, which depend from Claims 1 and 14, respectively), both of which are directed at using the results of a search. But the dependent claims support Xerox’s position, not Defendants’. “[A] claim in dependent form shall contain a reference to a claim previously set forth and then specify *a further limitation of the subject matter claimed.*” 35 U.S.C. § 112 (emphasis added). Since a dependent claim adds further limitations to the prior independent claim, the *en banc* Federal Circuit has stated that “the presence of a dependent claim that adds a particular limitation gives rise to a presumption that the limitation in question is *not present* in the independent claim”. *Phillips*, 415 F.3d at 1313-15 (emphasis added). Consequently, the references to search “results” in Claims 11 and 17 confirm that “a request for search results” is *not* required in Claim 1 (or in Claim 18 by parity of reasoning).

Defendants also point to language in column 49 of the specification that discusses “ranking” of results from a search. (Def. Br. at 15.) Not coincidentally, this is precisely what dependent Claim 11 covers, not Claims 1 and 18. Here again, Defendants ignore controlling patent law. The specification of the ’979 Patent describes a complete document management system, only a portion of which concerns queries based on document content. Furthermore, only discrete passages in that portion of the specification discuss query formulation. Binding precedent establishes that teachings in the specification that do not directly concern the particular invention defined by the claims of a particular patent may not be read into the claims of that patent as limitations. *See Broadcom Corp. v. Qualcomm Inc.*, 543 F.3d 683, 689 (Fed. Cir. 2008); *Ventana Med. Sys., Inc. v. Biogenex Labs., Inc.*, 473 F.3d 1173, 1181 (Fed. Cir. 2006).

For example, in *Broadcom v. Qualcomm*, the relevant claim was directed at the components of a digital signal processor (DSP). *Broadcom*, 543 F.3d at 688-89. Because the specification repeatedly discussed the interaction of the DSP with another component, called a “global controller”, the district court held that the claimed DSP necessarily required a global controller. *Id.* at 688. The Federal Circuit reversed, holding that because the global controller was the subject of other, unasserted patent claims, it could not be read into the asserted claims. *Id.* at 689. It noted that “each claim does not necessarily cover every feature disclosed in the specification. When the claim addresses only some of the features disclosed in the specification, it is improper to limit the claim to other, unclaimed features.” *Id.* (quoting *Ventana*, 473 F.3d at 1181). So, “[w]hile the specification does contemplate the possibility of a global controller, the claims of the [relevant] patent are directed solely to the DSP.” *Id.* Similarly, in the present case, simply because the specification contemplates the possibility of using a formulated query to request search results, Claims 1 and 18 do not require that the query be used only to request search results.<sup>3</sup> A formulated query could, for example, be stored or displayed.

As discussed in Xerox’s opening brief, the portions of the specification that actually teach query formulation repeatedly describe the query solely in terms of its contents. (X. Br. at 12-13.) That includes the sample query shown in Figure 39. Contrary to Defendants’ argument, the specification discussion of Figure 39 in column 50, lines 1-11 discusses the query solely in terms of its contents, not in terms of how a corresponding search would be performed or what results would be obtained. Indeed, the query of Figure 39 is not described as an “http request for results”, and the specification discussion does not say anything about sending the formulated query to a search engine to retrieve results.

---

<sup>3</sup> Defendants’ citation on page 15 of their brief to column 1, lines 37-39 of the specification is similarly misplaced, as that passage merely describes the overall system of the patent and does not purport to describe the specific invention of Claims 1 and 18.

All of the foregoing demonstrates why Xerox’s proposed construction of “query” as “a set of data specifying search criteria” best accords with the patent. Defendants’ attempt to construe “query” as a “request for search results” reflects a misconception of Claims 1 and 18.

**C. The Fallacy of Substituting “Confine” for “Restrict”**

The two preceding sections compel rejection of Defendants’ effort to engraft onto Claims 1 and 18 a requirement that a query must be a request for search results confined to a single category of information. The bulk of Defendants’ central claim construction contention is therefore manifestly incorrect. What remains is Defendants’ insistence that to formulate a query “to restrict a search” means only what is shown in a single specification figure—even though the law is clear that specification examples should not be read into claims, and even though the accompanying specification discussion conflicts with Defendants’ proposed construction.

Defendants devote much of their opening brief to attacking a strawman—the fiction that Xerox contends that the scope of a formulated query pursuant to Claims 1 or 18 can embrace *either* information concerning the set of entities *or* a category of information identified by a classification label. Defendants argue that such a query would have a broader scope than a query simply directed at a set of entities. Xerox has not argued for such a construction, as Xerox’s opening claim construction brief makes clear. Because the purpose of the claimed query formulation method is to improve the precision recall of information retrieval systems (*see* 979/48:37-39), adding classification label data to entity data in the formulated query must narrow the scope of any search governed by the query. The use of the word “restrict” in the claim makes this apparent.<sup>4</sup> The parties’ real disagreement concerns whether that narrowing must occur only in the manner taught in a single specification figure, Figure 39.

---

<sup>4</sup> To the extent any further clarification is required (and we don’t believe it is), Xerox would suggest the following amended construction: “The query includes data items corresponding to

Defendants argue that Figure 39 and the accompanying discussion in column 50, lines 1 to 11 support their position that a formulated query must invoke a search for information concerning an entity only in a single category of information. (Def. Br. at 8-9.) By attempting to read a single example into the claims as a limitation, this argument ignores basic principles of claim construction. As this Court has recognized, “unless the patentee has demonstrated a clear intention to limit the claim scope using words or expressions of manifest exclusion or restriction,” a single specification embodiment cannot be used to limit the claims. *Wyeth*, 2011 WL 1043575, at \*3 (quoting *Liebel-Flarsheim*, 358 F.3d at 906). No such words or expressions appear in the '979 Patent.

Figure 39 shows a sample query formulated using the method of Claims 1 and 18. The query contains the set of entities “seven up” and is directed to a particular category of documents in the Google Web Directory (the current version of which may be accessed at <http://www.google.com/dirhp>). However, the patent specifically notes that this is only an example: “FIG. 39 illustrates an *example* of a query 3930 contextualized using classification labels 3920 . . . .” (979/50:1-2 (emphasis added).) In turn, the notation “e.g.” precedes the actual sample query in the box labeled 3930 in Figure 39, and the '979 Patent specification makes it clear that the claims are intended to be broader than the specification embodiments:

The invention has been described with reference to particular embodiments. Modifications and alterations will occur to others upon reading and understanding this specification taken together with the drawings. The embodiments are but examples, and various alternatives, modifications, variations or improvements may be made by those skilled in the art from this teaching which are intended to be encompassed by the following claims[.] (979/76:1-8.)

---

one or more entities identified in the ‘automatically identifying’ step and one or more classification labels assigned in the ‘automatically categorizing’ step. The query’s scope is narrowed by the classification labels.”

Because there is no teaching that the claims are limited to the precise form of query narrowing shown in Figure 39, this single specification example cannot be read into the claims.

Furthermore, when quoting from the patent's teachings concerning Figure 39, Defendants distort their meaning with selective underlining: "[t]he specification explains that 'the search is focused on documents found in the single node of the document hierarchy genetics, at 3910'." (Def. Br. at 8 (quoting 979/50:10-11) (emphasis in original).) Defendants' selective emphasis seeks to elide the fact that the passage quoted above expressly teaches that a search need only be "*focused*" on documents in the category node "genetics"—not that a search in accordance with the invention can return *only* documents in the node or category of genetics. An example of a search that is focused on a category, but not limited solely to documents that fall within that category, would be a search within a primary category and in other categories that gives greater weight to results from the primary category.<sup>5</sup> And, as noted above, the specification specifically states that more than one classification label may be assigned to a given piece of document content. (979/49:31-37.) Accordingly, as long as the inclusion of classification labels in a formulated query narrows and focuses any ensuing search, the claim language is satisfied. Defendants' attempt to read in a far more limited construction should be rejected.

Defendants' use of the word "confine" is, in any event, unsupported. Defendants rely principally on extrinsic evidence—definitions from two general dictionaries—to replace the claim term "restrict" with "confine". Extrinsic evidence is hardly the place to start and, of course, carries little weight in claim construction. *Phillips*, 415 F.3d at 1318-19. Moreover, the Federal Circuit has taken a cautious view toward use of general dictionaries in claim construction

---

<sup>5</sup> See, e.g., Transcript of G. Grefenstette Deposition at 11:18-12:9, *Xerox Corp. v. Google, Inc.*, Civ. A. No. 10-136 (D. Del. Mar. 30, 2011), attached as Ex. 2; Transcript of J. Shanahan Deposition at 192:18-193:17, *Xerox Corp. v. Google, Inc.*, Civ. A. No. 10-136 (D. Del. Mar. 30, 2011), attached as Ex. 3 (discussing a query "focused" on a node, or category).

(*id.* at 1321-22), and Defendants’ argument illustrates why. Although Defendants cite two dictionaries as supporting their rewrite of the claim language (*see* Def. Br. at 8), other dictionaries give different definitions. For example, the *Cambridge Dictionary Online* defines “restrict” as “to limit (an intended action) esp. by setting the conditions under which it is allowed to happen”. (Ex. 4.) And *Webster’s Third New International Dictionary* defines “restrict” as

**1** : to set bounds or limits to : hold within bounds: as **a** : to check free activity, motion, progress, or departure of : RESTRAIN... **b** : to check, bound, or decrease the range, scope, or incidence of : set what is to be included or embraced by : bar or carefully govern addition or increment to.... (Ex. 5.)

There is no principled basis for selecting some extrinsic general definitions of “restrict” over others, and no sound reason to replace “restrict” (the patent’s term) with “confine” (a term not used in the patent).

Defendants also cite various passages from the specification that merely echo the claim language and repeat the claim term “restrict”. These citations provide no support for Defendants’ narrow construction of “restrict” and certainly no support for swapping “restrict” for “confine”.

Defendants’ reliance on the prosecution history is equally misplaced. Defendants make the point that Xerox told the Patent Office that the scope of the claimed query is restricted, not broadened, by adding classification labels, thereby disclaiming any claim interpretation<sup>6</sup> that a query including both entities and classification labels would have a broader scope than one with only entities. True—but Xerox is not arguing, and has never argued, for the strawman position Defendants attack. Moreover, “[i]n evaluating whether a patentee has disavowed claim scope, context matters”. *i4i Ltd. P’ship v. Microsoft Corp.*, 598 F.3d 831, 842-43 (Fed. Cir.

---

<sup>6</sup> “The purpose of consulting the prosecution history in construing a claim is to exclude any interpretation that was disclaimed during prosecution.” *Phillips*, 415 F.3d at 1317 (quoting *Chimie v. PPG Indus., Inc.*, 402 F.3d 1371, 1384 (Fed. Cir. 2005)).

2010) (rejecting attempts to limit claim scope on the basis of statements plucked from prosecution history). In distinguishing the Rennison prior art patent during prosecution, Xerox repeatedly stressed the novel combination of Claim 1 as a whole. (*See, e.g.*, Ex. 6 at 5-6, 8.) In discussing the novelty of that combination, Xerox simply repeated the claim language, including using the claim term “restrict”. Xerox never argued that any particular form of query restriction was required by the patent, much less that a query must invoke a search only within a single category of information. Indeed, in allowing the ’979 Patent, the examiner agreed that Claim 1 was novel as a whole; he never indicated that allowance was based in any way on the narrow reading of “restrict” proffered by Defendants. (*See* Ex. 7.) Accordingly, pursuant to *Phillips*, the prosecution history does not support Defendants’ construction.

\* \* \*

Defendants’ overarching effort to constrict the scope of claims 1 and 18 by requiring that the query only request search results confined to a single category of information fails for all the reasons noted above. Xerox’s constructions should be accepted, and Defendants’ proposed constructions should be rejected.

#### **IV. THE FORMAT-OF-THE-LABEL FALLACY**

Defendants’ meritless arguments for construing the term “classification label” as a “classifying word or phrase” contravene Federal Circuit precedent emphasizing that claim terms are interpreted from the perspective of one skilled in the relevant field and that a specification example should not be read into the claims.

Defendants begin by arguing that the plain meaning of “label” is a word or phrase, citing snippets from two general-purpose dictionaries for that proposition. However, the Federal Circuit has emphasized *en banc* that words in patent claims are interpreted from a technical perspective and cautioned against the use of general dictionary definitions. *Phillips*, 415 F.3d at

1321-22. Here, the definitions that Defendants rely on to support their position are plainly inapposite. For example, in quoting a portion of definition (c) of “label” from a dictionary, Defendants fail to mention that the dictionary goes on to provide three contexts for that definition, none of them relevant to the ’979 Patent:

book) <read the ~ on the bottle> **c** : a descriptive, classifying, or identifying word or phrase: as (1) : EPITHET (the term stream of consciousness . . . is already established as a literary ~ —Robert Humphrey) <acquired the ~ of “playboy” which seemed to stick —Brian Crozier> <hanging the subversive ~ on their own liberal clergy —Ralph Winnett> (2) : a word or phrase used with but not as part of a dictionary definition usu. in abbreviated form and distinctive type to provide information (as grammatical function or area or level of usage) about the word defined (the ~ *obsolete* is abbreviated *obs*) (3) : a newspaper headline merely identifying the subject matter of an article rather than summarizing action **6** : a

(*Webster’s Third New Int’l Dictionary, Unabridged* (2002), 3/29/11 Roberts Decl., D.I. 143,

Ex. 4.) The ’979 Patent obviously does not use “label” (1) as an epithet, (2) in the context of a dictionary entry, or (3) as a newspaper headline. Defendants then cite definition 3 from another dictionary:

a person, group, intellectual movement, etc. **3.** a word or phrase indicating that what follows belongs in a particular category or classification: *The following definition has the label “Archit.”* **4.** *Archit.* a molding or dripstone over a door or window, esp. one that extends horizontally across the top of the opening and vertically downward for a certain distance at the sides. **5.** a brand

(*Random House Unabridged Dictionary* (1993), 3/29/11 Roberts Decl., D.I. 143, Ex. 1.) This definition reflects the same irrelevant sense of “label” discussed in definition (c)(2) from the *Webster’s* dictionary. The ’979 Patent is not directed at dictionary entries. Thus, Defendants’ general dictionary definitions have no bearing on the claims of the ’979 Patent.

In contrast to such inapposite general definitions, relevant technical dictionaries make clear that a “label” is simply an identifier and not necessarily a word or a phrase. The *IBM Dictionary of Computing* (1994) defines “label” as “[a]n identifier within or attached to a set of data elements”. (X. Br. at 7.) The *McGraw-Hill Computer Desktop Encyclopedia* (2001)

defines label in the context of data management “a made-up name that is assigned to a file, field or other data structure” (Ex. 8), while the *McGraw-Hill Dictionary of Computing & Communications* (2003) defines label as “[a] data item that serves to identify a data record (much in the same way as a key is used) . . . .” (Ex. 9.) None of these technical definitions require a label to be a word or a phrase.

Defendants further contend that because the specification uses examples of words as classification labels, the claimed “classification label” can only be a word. But this is precisely the claim construction methodology condemned by the Federal Circuit:

[A]lthough the specification often describes very specific embodiments of the invention, we have repeatedly warned against confining the claims to those embodiments. In particular, we have expressly rejected the contention that if a patent describes only a single embodiment, the claims of the patent must be construed as being limited to that embodiment. That is not just because section 112 of the Patent Act requires that the claims themselves set forth the limits of the patent grant, but also because persons of ordinary skill in the art rarely would confine their definitions of terms to the exact representations depicted in the embodiments. *Phillips*, 415 F.3d at 1323 (citations omitted).

Because the technical meaning of “label” in computer science is simply an identifier, and because words can be identifiers, it is neither surprising nor significant that the specification uses words (such as DMOZ category names) as examples of labels. But there is no teaching in the specification that it is necessary or desirable to depart from the ordinary technical meaning of “label” and use *only* words or phrases as classification labels. *See Wyeth*, 2011 WL 1043575, at \*3. To the contrary, as discussed in Xerox’s opening brief (at 8), the specification teaches that the label is simply a variable and places no restriction on its format. (*See* 979/43:41-56.)

Nor would a teaching that a classification label must be a word make any sense. The ’979 Patent describes and claims computer-implemented inventions. (*See* 979/74:30-75:67.) In computers, all data is represented by numbers. Consequently, even if document content were categorized into the category of “genetics,” and the associated classification label happened to be

the word “genetics”, it would nevertheless exist in the computer in numerical form. For example, if the computer system uses ASCII codes to represent letters, the word “genetics” would be expressed as “01100111 01100101 01101110 01100101 01110100 01101001 01100011 01110011”. (*See* Ex. 10, at 38-40, 597.) Hence, a person of ordinary skill in computer science would see no reason to require that a classification label be a word or a phrase, as opposed to a number, a mix of numbers or letters, a variable, or any other suitable identifier.

As for Defendants’ proposed requirement that a classification label be “classifying,” which they do not explain in their opening brief, Defendants here confuse the role of the classification label and the categories (“classes”) that make up the “organized classification of document content” (“OCDC”) of Claims 1 and 18. It is the categories of an OCDC such as DMOZ that are used for classification. (*See* 979/41:53-60.) Pursuant to Step (a) of Claim 1, classification labels are simply associated with those OCDC categories; labels are then assigned to document content as the result *of* classification, as opposed to being used *for* classification. Thus, in Step (c), when categorization of document content outputs specific OCDC categories, the associated labels are assigned and used for query formulation. (*See, e.g.*, 979/49:31-37.) Accordingly, Defendants’ characterization of classification labels as “a *classifying* word or phrase” should be rejected along with the rest of their proposed construction.

In contrast to Defendants’ proposed construction, Xerox’s construction accurately reflects the relationship between OCDC classes and classification labels in Claim 1, which expressly associates labels with classes, while making it clear that labels can be in any format. Contrary to Defendants’ criticism, this is not mere rearrangement of claim language. Xerox’s construction defines “label” by both its function (to identify an OCDC class) and its potential format (any format). Accordingly, the Court should reject Defendants’ proposed construction and adopt Xerox’s construction.

**V. DEFENDANTS FAIL TO SHOW THAT THE TERM “SELECTED DOCUMENT CONTENT” IS INDEFINITE**

As discussed in Xerox’s opening brief, the meaning of “selected document content” is in no way ambiguous or indefinite. (*See* X. Br. at 3-7.) The language of Claim 1 indicates that “selected document content” is the input to the claimed method: it is the document content in which entities (for example, names or places) are identified and the document content that is categorized. The portion of the specification that teaches query formulation plainly indicates that this content can comprise all or part of a document. (979/48:52-55.) Additionally, because the claimed method is computer-implemented, the content must be in electronic form.

In their brief, Defendants nowhere mention the stringent Federal Circuit tests governing indefiniteness. For example, “[a] claim will be found indefinite only if it is insolubly ambiguous, and no narrowing construction can properly be adopted . . . .” *Praxair, Inc. v. ATMI, Inc.*, 543 F.3d 1306, 1319 (Fed. Cir. 2008) (citation omitted). Instead of mentioning these tests, Defendants merely assert that “selected document content” has no antecedent basis and is therefore indefinite. Defendants’ argument is baseless. The sole case they cite for that proposition, *Halliburton Energy Servs., Inc. v. M-I LLC*, 514 F.3d 1244, 1249 (Fed. Cir. 2008), does not apply the antecedent basis test. Instead, in the course of emphasizing the exacting nature of the indefiniteness inquiry, *Halliburton* merely lists lack of antecedent basis as a potential ground of indefiniteness, citing *Energizer Holdings, Inc. v. Int’l Trade Comm’n*, 435 F.3d 1366, 1370-71 (Fed. Cir. 2006). *Energizer*, in turn, clearly explains that indefiniteness for lack of antecedent basis occurs when a claim term references a non-existent earlier claim term. 435 F.3d at 1370-71. For example, indefiniteness for lack of antecedent basis occurs “where a claim refers to ‘said lever’ or ‘the lever,’ where the claim contains no earlier recitation or

limitation of a lever and where it would be unclear as to what element the limitation was making reference”. MPEP § 2173.05(e) (Ex. 1). No such situation is present here.

Defendants also rely on the red-herring argument that “selected document content” is indefinite because the intrinsic evidence “is silent as to how document content is selected or by whom or what it may be selected”. (Def. Br. at 16.) But which particular document content is used for entity identification and categorization, including the manner in which that content is chosen or selected, is totally irrelevant to the method of Claims 1 and 18. The claims are not directed at how document content is selected for use in the claimed method, and the claimed method can be performed regardless of who or what performs the selection.

As for what content can be used to perform the claimed method, the specification section that specifically discusses entity identification and categorization in the context of query formulation is crystal-clear that this content can consist of all or part of a document:

In operation as shown in FIG. 38, the document content 3612 or alternatively limited context (i.e., words, sentences, or paragraphs) surrounding the entity 3808 is analyzed by categorizer 3610 to produce a set of categories 3620. (979/48:52-55.)

Since the relevant portion of the specification readily explains what is intended by the term “selected document content”, all of Defendants’ citations to other portions of the specification that have nothing to do with query formulation are irrelevant.

As the Federal Circuit held in *Halliburton*, a claim is indefinite only if “an accused infringer shows by clear and convincing evidence that a skilled artisan could not discern the boundaries of the claim based on the claim language, the specification, and the prosecution history, as well as her knowledge of the relevant art area”. 514 F.3d at 1249-50. Defendants have not come close to meeting this exacting standard.

**VI. DEFENDANTS' REMAINING CLAIM CONSTRUCTION ARGUMENTS ARE WITHOUT MERIT**

**A. "Characteristic Vocabulary" Describes a Class in the Organized Classification of Document Content**

Claim 10 requires that each class in the "organized classification of document content" ("OCDC") of Step (a) of Claim 1 have an associated "characteristic vocabulary." Despite that express association, Defendants argue that the characteristic vocabulary does not describe an OCDC class, but rather a "category," which, Defendants have indicated, means the "category of information" in an information retrieval system. (Def. Br. at 20; Ex. 11, at 1-2.) Defendants' sole support for this construction is a sentence in the specification that states that "category vocabulary"—which the specification uses as a synonym for "characteristic vocabulary" (979/49:44-45)—"consists of one or more terms that describe the category". (979/51:36-37.) However, the specification uses "category" as a synonym for "class," as in an OCDC class. (*See* 979/41:53-56 ("The goal of a text classification system... is to classify a document 3612 into a set of one or more classes 3620, which are also referred to as categories.")) As demonstrated in detail in Xerox's opening brief (at 16-17), when the specification states that the "category vocabulary" describes a "category," it is referring to the classes/categories used for categorization, *i.e.*, the classes/categories that comprise the OCDC. Indeed, the specification uses the same reference number (3620) to refer to a class in the OCDC and the category described by category vocabulary. (*See, e.g.*, 979/41:53-56; 979/51:34-37.) Because the intrinsic evidence directly contradicts Defendants' proposed construction, it should be rejected.

**B. Nothing in Claims 1 and 18 or the Specification Requires the "Automatically Identifying" and "Automatically Categorizing" Steps To Be Fully Completed before the "Automatically Formulating the Query" Step Is Commenced**

"[A]s a general rule the claim is not limited to performance of the steps in the order recited, unless the claim explicitly or implicitly *requires* a specific order." *Baldwin Graphic*

*Sys., Inc. v. Seibert, Inc.*, 512 F.3d 1338, 1345 (Fed. Cir. 2008) (emphasis added). Defendants argue that because Step (d) (“automatically formulating the query”) of Claim 1 utilizes the entities and classification labels respectively derived in Steps (b) and (c), those steps must occur before Step (d). (Def. Br. at 18-19.) But Defendants provide no support whatsoever for their contention that the claim language somehow requires Step (b) (“automatically identifying a set of entities”) and Step (c) (“automatically categorizing”) of Claim 1 to be *fully completed* before Step (d) is even begun.

Although Defendants imply that it is Xerox’s burden to show that the claims should not be read restrictively, the law is otherwise. Unless the logic or grammar of the claim language, or the content of the specification, compels a contrary result, “the sequence in which such steps are written is not a requirement”. *Altiris, Inc. v. Symantec Corp.*, 318 F.3d 1363, 1369-70 (Fed. Cir. 2003) (citations omitted). This principle includes whether one claim step must be completed before another is begun. *See Bell Comms. Research, Inc. v. Fore Sys., Inc.*, 62 Fed. Appx. 951, 954-55 (Fed. Cir. 2003) (non-precedential) (“[t]he precise question here is not whether the first step must be performed before the second step is performed, but whether it must be completed before the second step is begun”; also citing the *Altiris* test as controlling for this question).

Nothing in the logic or grammar of Claim 1 requires that entity identification or categorization must be completed before query formulation is begun, and nothing in the specification teaches such a requirement. To the contrary, the specification explicitly contemplates that the claimed method can be implemented using parallel processors or multiple computers. (979/75:15-22.) In such systems, a person of ordinary skill in computer science could readily implement the method of Claim 1 by having one computer or processor handle entity identification, a second perform categorization of document content, and a third work on formulation of a query, inserting results of the entity extraction and the categorization processes

as they become available. And nothing in the claims or the specification precludes other possible embodiments, for example, storing a partially formulated query, and then later adding more entity data and classification labels to it. In sum, because nothing in the claims or the specification requires that Steps (b) and (c) be completed before Step (d) begins, there is no basis for reading such a requirement into the claims under the law governing order of steps.

**C. Nothing in the Claims or the Specification Requires the Steps of Claims 1 and 18 To Be Fully Completed before the Steps of Claims 2 and 19 Are Commenced**

Although Defendants argue that the language of the claims requires that the step of Claim 2 be performed only after a query is formulated in Claim 1, the claims say nothing of the sort. Claim 2 simply covers “limiting the query by adding terms . . .”. (979/76:33.) Nothing in the claim language requires adding limiting terms only *after* a query has been formulated in accordance with Claim 1. Claim 2 does not say, as Defendants would read it, “limiting said *fully formulated* query by adding terms”. As a matter of grammar, logic and common sense, it is equally possible to add terms *during* query formulation. Indeed, the specification explicitly discloses an embodiment in which the classification labels of Step (c) of Claim 1 are “appended” to the query *after* it contains both entities identified in Step (b) of Claim 1 and the “terms” (*i.e.*, aspect vector) of Claim 2. (*See* X. Br. at 18-19.) Because Defendants’ proposed construction is contradicted by the specification, it must be rejected. *See Wyeth*, 2011 WL 1043575, at \*4 (“a claim interpretation that would exclude the inventor’s device is rarely the correct interpretation”) (quoting *Osram GmbH v. Int’l Trade Comm’n*, 505 F.3d 1351, 1358 (Fed. Cir. 2007)).

**VII. CONCLUSION**

For the foregoing reasons, the Court should adopt Xerox’s proposed constructions of the disputed claim terms and phrases.

ASHBY & GEDDES,

*/s/ John G. Day*

---

Lawrence C. Ashby (I.D. #468)  
John G. Day (I.D. #2403)  
Lauren E. Maguire (I.D. #4261)  
500 Delaware Avenue, 8th Floor  
P.O. Box 1150  
Wilmington, DE 19899  
(302) 654-1888  
lashby@ashby-geddes.com  
jday@ashby-geddes.com  
lmauire@ashby-geddes.com

*Attorneys for Plaintiff Xerox Corporation*

*Of Counsel:*

Richard J. Stark  
Andrei Harasymiak  
CRAVATH, SWAINE & MOORE LLP  
Worldwide Plaza  
825 Eighth Avenue  
New York, NY 10019  
(212) 474-1000

Dated: April 26, 2011