

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

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

ALBERTO R. GONZALES, in his official
capacity as Attorney General of the United
States,

NO. CV 06-8006MISC JW

Plaintiff,

**ORDER GRANTING IN PART AND
DENYING IN PART MOTION TO
COMPEL COMPLIANCE WITH
SUBPOENA DUCES TECUM**

v.

GOOGLE, INC.,

Defendant.

I. INTRODUCTION

This case raises three vital interests: (1) the national interest in a judicial system to reach informed decisions through the power of a subpoena to compel a third party to produce relevant information; (2) the third-party's interest in not being compelled by a subpoena to reveal confidential business information and devote resources to a distant litigation; and (3) the interest of individuals in freedom from general surveillance by the Government of their use of the Internet or other communications media.

In aid of the Government's position in the case of ACLU v. Gonzales, Civil Action No. 98-CV-5591 pending in the Eastern District of Pennsylvania, United States Attorney General Alberto R. Gonzales has subpoenaed Google, Inc., ("Google") to compile and produce a massive amount of information from Google's search index, and to turn over a significant number of search queries entered by Google users. Google timely objected to the Government's request. Following the

1 requisite meet and confer, the Government filed the present Miscellaneous Action in this District to
2 compel Google to comply with the subpoena. On March 14, 2006, this Court held a hearing on the
3 Government's Motion.¹ At that hearing, the Government made a significantly scaled-down request
4 from the information it originally sought. For the reasons explained in this Order, the motion to
5 compel, as modified, is GRANTED as to the sample of URLs from Google search index and
6 DENIED as to the sample of users' search queries from Google's query log.

7
8 II. PROCEDURAL BACKGROUND

9 In 1998, Congress enacted the Child Online Protection Act ("COPA"), which is now codified
10 as 47 U.S.C. § 231. COPA prohibits the knowing making of a communication by means of the
11 World Wide Web, "for commercial purposes that is available to any minor and that includes material
12 that is harmful to minors," subject to certain affirmative defenses. 47 U.S.C. § 231 (a)(1). For this
13 purpose, the statute defines the phrase "material that is harmful to minors" to mean material that is
14 either obscene or material that meets each prong of a three-part test: "(A) the average person,
15 applying contemporary community standards, would find, taking the material as a whole and with
16 respect to minors, is designed to appeal to, or is designed to pander to, the prurient interest; (B)
17 depicts, describes, or represents, in a manner patently offensive with respect to minors, an actual or
18 simulated sexual act or sexual conduct, an actual or simulated normal or perverted sexual act, or a
19 lewd exhibition of the genitals or post-pubescent female breast; and (C) taken as a whole, lacks
20 serious literary, artistic, political, or scientific value for minors." 47 U.S.C. § 231 (e)(6).

21 Upon enactment of COPA, the American Civil Liberties Union and several other plaintiffs
22 ("Plaintiffs") filed an action in the Eastern District of Pennsylvania, challenging the constitutionality
23 of the Act. The district court granted Plaintiffs' motion for a preliminary injunction on the grounds
24 that COPA is likely to be found unconstitutional on its face for violating the First Amendment rights
25 of adults. ACLU v. Reno, 31 F. Supp. 2d 473 (E.D. Pa. 1998). The United States Court of Appeals

26
27 ¹ The Court continued the hearing date originally proposed by the parties in order to allow
28 for amici to prepare and submit their briefs to the Court.

1 for the Third Circuit affirmed the grant of the preliminary injunction. ACLU v. Reno, 217 F.3d 162
2 (3d Cir. 2000). After granting certiorari, the Supreme Court of the United States vacated the
3 judgment of the Third Circuit, and remanded the case to that court for further review of the district
4 court's grant of preliminary injunction in favor of Plaintiffs. The Third Circuit again affirmed the
5 preliminary injunction, ACLU v. Ashcroft, 322 F.3d 240 (3d Cir. 2003), and the Supreme Court
6 again granted certiorari.

7 The Supreme Court affirmed the preliminary injunction and held that there was an
8 insufficient record before it by which the Government could carry its burden to show that less
9 restrictive alternatives may be more effective than the provisions of COPA. Ashcroft v. ACLU, 542
10 U.S. 656, 673 (2004). Of these alternatives directed at preventing minors from viewing "harmful to
11 minors" material on the Internet, the Court focused on blocking and filtering software programs
12 which "impose selective restrictions on speech at the receiving end, not universal restrictions at the
13 source." Id. at 667. To "allow the parties to update and supplement the factual record to reflect
14 current technological realities," the Court remanded the case for a trial on the merits. Id. at 672.

15 Following remand, Plaintiffs filed a First Amended Complaint ("FAC"). (98-CV-5591LR,
16 E.D. Pa., Docket Item No. 175). Apparently, in preparing its defense, the Government initiated a
17 study designed to somehow test the effectiveness of blocking and filtering software. To provide it
18 with data for its study, the Government served a subpoena on Google, America Online, Inc.
19 ("AOL"), Yahoo! Inc. ("Yahoo"), and Microsoft, Inc. ("Microsoft"). The subpoena required that
20 these companies produce a designated listing of the URLs which would be available to a user of
21 their services. The subpoena also required the companies to produce the text of users' search
22 queries. AOL, Yahoo, and Microsoft appear to be producing data pursuant to the Government's
23 request. Google, however, objected.

24 Google is a Delaware corporation headquartered in Mountain View, CA, that, like AOL,
25 Yahoo, and Microsoft, also provides search engine capabilities. Based on the Government's
26 estimation, and uncontested by Google, Google's search engine is the most widely used search
27

1 engine in the world, with a market share of about 45%. The search engine at Google yields URLs in
2 response to a search query entered by a user. The search queries entered may be of varying lengths,
3 and incorporate a number of terms and connectors. Upon receiving a search query, Google produces
4 a responsive list of URLs from its search index in a particular order based on algorithms proprietary
5 to Google.

6 The initial subpoena to Google sought production of an electronic file containing two general
7 categories. First, the subpoena requested "[a]ll URL's that are available to be located to a query on
8 your company's search engine as of July 31, 2005." (Decl. of Joel McElvain, Ex. A ("Subpoena") at
9 4.) In negotiations with Google, this request was later narrowed to a "multi-stage random" sampling
10 of one million URLs in Google's indexed database. As represented to the Court at oral argument,
11 the Government now seeks only 50,000 URLs from Google's search index. Second, the government
12 also initially sought "[a]ll queries that have been entered on your company's search engine between
13 June 1, 2005 and July 31, 2005 inclusive." (Subpoena at 4.) Following further negotiations with
14 Google, the Government narrowed this request to all queries that have been entered on the Google
15 search engine during a one-week period. During the course of the present Miscellaneous Action, the
16 Government further restricted the scope of its request, and now represents that it only requires 5,000
17 entries from Google's query log in order to meet its discovery needs.

18 Despite these modifications in the scope of the subpoena, Google maintained its objection to
19 the Government's requests. Before the Court is a motion to compel Google to comply with the
20 modified subpoena, namely, for a sample of 50,000 URLs from Google's search index and 5,000
21 search queries entered by Google's users from Google's query log.

22 III. STANDARDS

23 Rule 45 of the Federal Rules of Civil Procedure governs discovery of nonparties by
24 subpoena. FED. R. CIV. P. 45 ("Rule 45"). The Advisory Committee Notes to the 1970 Amendment
25 to Rule 45 state that the "scope of discovery through a subpoena is the same as that applicable to
26 Rule 34 and other discovery rules." Rule 45 advisory committee's note (1970). Under Rule 34, the
27

1 rule governing the production of documents between parties, the proper scope of discovery is as
2 specified in Rule 26(b). FED. R. CIV. P. 34. See also Heat & Control, Inc. v. Hester Industries, Inc.,
3 785 F.2d 1017 (Fed. Cir. 1986) ("rule 45(b)(1) must be read in light of Rule 26(b)"); Exxon
4 Shipping Co. v. U.S. Dept. of Interior, 34 F.3d 774, 779 (9th Cir. 1994) (applying both Rule 26 and
5 Rule 45 standards to rule on a motion to quash subpoena).

6 Rule 26(b), in turn, permits the discovery of any non-privileged material "relevant to the
7 claim or defense of any party," where "relevant information need not be admissible at trial if the
8 discovery appears reasonably calculated to lead to the discovery of admissible evidence." Rule
9 26(b)(1). Relevancy, for the purposes of discovery, is defined broadly, although it is not without
10 "ultimate and necessary boundaries." Pacific Gas and Elec., Co. v. Lynch, No. C-01-3023 VRW,
11 2002 WL 32812098, at *1 (N.D. Cal. August 19, 2002) (citing Hickman v. Taylor, 329 U.S. 495,
12 507 (1947)).

13 Rule 26 also specifies that "[a]ll discovery is subject to the limitations imposed by Rule
14 26(b)(2)(i), (ii), and (iii)" which requires that discovery methods be limited where:

15 (i) the discovery sought is unreasonably cumulative or duplicative, or is obtainable
16 from some source that is more convenient, less burdensome, or less expensive; (ii)
17 the party seeking discovery has had ample opportunity by discovery in the action to
18 obtain the information sought; or (iii) the burden or expense of the proposed
19 discovery outweighs its likely benefit, taking into account the needs of the case, the
20 amount in controversy, the parties' resources, the importance of the issues at stake
21 in the litigation, and the importance of the proposed discovery in resolving the
22 issues.

23 The Advisory Committee Notes to the 1983 amendments to Rule 26 state that "[t]he objective is to
24 guard against redundant or disproportionate discovery by giving the court authority to reduce the
25 amount of discovery that may be directed to matters that are otherwise proper subjects of inquiry."

26 However, the commentators also caution that "the court must be careful not to deprive a party of
27 discovery that is reasonably necessary to afford a fair opportunity to defend and prepare the case."

28 Rule 26 advisory committee's note (1983).

In addition to the discovery standards under Rule 26 incorporated by Rule 45, Rule 45 itself
provides that "on timely motion, the court by which a subpoena was issued shall quash or modify the

1 subpoena if it...subjects a person to undue burden." Rule 45(3)(A). Of course, "if the sought-after
2 documents are not relevant, nor calculated to lead to the discovery of admissible evidence, then any
3 burden whatsoever imposed would be by definition 'undue.'" Compaq Computer Corp. v. Packard
4 Bell Elec., Inc., 163 F.R.D. 329, 335-36 (N.D. Cal. 1995). Underlying the protections of Rule 45 is
5 the recognition that "the word 'non-party' serves as a constant reminder of the reasons for the
6 limitations that characterize 'third-party' discovery." Dart Indus. Co. v. Westwood Chem. Co., 649
7 F.2d 646, 649 (9th Cir. 1980) (citations omitted). Thus, a court determining the propriety of a
8 subpoena balances the relevance of the discovery sought, the requesting party's need, and the
9 potential hardship to the party subject to the subpoena. Heat & Control, 785 F.2d at 1024.

10 IV. DISCUSSION

11 Google primarily argues that the information sought by the subpoena is not reasonably
12 calculated to lead to evidence admissible in the underlying litigation, and that the production of
13 information is unduly burdensome. The Court discusses each of these objections in turn, as well as
14 the Court's own concerns about the potential interests of Google's users.

15 A. Relevance

16 Any information sought by means of a subpoena must be relevant to the claims and defenses
17 in the underlying case. More precisely, the information sought must be "reasonably calculated to
18 lead to admissible evidence." Rule 26(b). This requirement is liberally construed to permit the
19 discovery of information which ultimately may not be admissible at trial. Overbroad subpoenas
20 seeking irrelevant information may be quashed or modified. See, e.g., Moon v. SCP Pool Corp., 232
21 F.R.D. 633, 637 (C.D. Cal. 2005) (quashing subpoena seeking the production of all purchasing
22 information where the underlying contract dispute was limited to a particular geographic region);
23 W.E. Green v. Baca, 219 F.R.D. 485, 490 (C.D. Cal. 2003) (providing a survey of cases where in
24 limiting the scope of a subpoena, district courts "effectively sustain[] an objection that the requests
25 are vague, ambiguous, or overbroad in part, and overrules in part").

26 This Court does not have the benefit of involvement with the underlying litigation. The
27
28

1 Court adheres to the principle stated in Truswal Systems Corp. v. Hydro-Air Engineering, Inc., 813
2 F.2d 1207, 1211-12 (Fed. Cir. 1987): "A district court whose only connection with a case is
3 supervision of discovery ancillary to an action in another district should be especially hesitant to
4 pass judgment on what constitutes relevant evidence thereunder. Where relevance is in doubt . . . the
5 court should be permissive."

6 However, the Court does not construe a general policy of permissiveness to require this
7 Court to abdicate its responsibility to review a subpoena under the Federal Rules when presented
8 with a motion to compel. The Court has reviewed the decisions comprising the lengthy procedural
9 history of this case in the Eastern District of Pennsylvania, the Third Circuit, and the Supreme Court,
10 as well as Plaintiffs' current complaint. The Court has heard the parties at oral argument² and
11 proceeds to consider the merits of the Government's motion.

12 1. Sample of URLs

13 As narrowed by negotiations with Google and through the course of this Miscellaneous
14 Action, the Government now seeks a sample of 50,000 URLs from Google's search index. In
15 determining whether the information sought is reasonably calculated to lead to admissible evidence,
16 the party seeking the information must first provide the Court with its plans for the requested
17 information. See Northwestern Memorial v. Ashcroft, 362 F.3d 923, 931 (7th Cir. 2004). The
18 Government's disclosure of its plans for the sample of URLs is incomplete. The actual methodology
19 disclosed in the Government's papers as to the search index sample is, in its entirety, as follows: "A
20 human being will browse a random sample of 5,000-10,000 URLs from Google's index and
21 categorize those sites by content" (Supp. Decl. of Phillip B. Stark, Ph.D ("Supp. Stark Decl.") ¶ 4)
22 and from this information, the Government intends to "estimate...the aggregate properties of the
23 websites that search engines have indexed." (Government's Reply Memorandum in Support of the
24 Motion to Compel Compliance with Subpoena Duces Tecum ("Reply"), Docket Item No. 21 at 4:8-

26 ² Counsel for Plaintiffs also appeared at the Court's hearing on the Government's Motion to
27 Compel.

1 9.) The Government's disclosure only describes its methodology for a study to categorize the URLs
2 in Google's search index, and does not disclose a study regarding the effectiveness of filtering
3 software. Absent any explanation of how the "aggregate properties" of material on the Internet is
4 germane to the underlying litigation³, the Government's disclosure as to its planned categorization
5 study is not particularly helpful in determining whether the sample of Google's search index sought
6 is reasonably calculated to lead to admissible evidence in the underlying litigation.

7 Based on the Government's statement that this information is to act as a "test set for the
8 study" (Reply at 3:20) and a general statement that the purpose of the study is to "evaluate the
9 effectiveness of content filtering software," (Reply at 3:2-5) the Court is able to envision a study
10 whereby a sample of 50,000 URLs from the Google search index may be reasonably calculated to
11 lead to admissible evidence on measuring the effectiveness of filtering software. In such a study, the
12 Court imagines, the URLs would be categorized, run through the filtering software, and the
13 effectiveness of the filtering software ascertained as to the various categories of URLs. The
14 Government does not even provide this rudimentary level of general detail as to what it intends to do
15 with the sample of URLs to evaluate the effectiveness of filtering software, and at the hearing
16 neither confirmed nor denied the Court's speculations about the study.⁴ In fact, the Government
17 seems to indicate that such a study is not what it has in mind: "[t]he government seeks this

18
19
20 ³ Whether adult material exists on the Internet could not seriously be contested by Plaintiffs
21 with web content describing the slang terms "teabagging" and "pearl necklace" in graphic detail
22 (FAC at 43), or websites which contain "numerous photographs of nude men and women in sexual
23 poses with one another, and erotic stories that include graphic sexual scenes" (FAC at 34). Such a
24 reading of the Complaint is also supported by the narrow question posed by the Supreme Court to be
25 answered on remand for trial on the merits.

26 ⁴ The lack of disclosure on the part of the government is particularly striking when seen in
27 the context of the time that the Government has had to prepare this issue. The Supreme Court's
28 directive to the Government to address the effectiveness of filtering software was issued in 2004.
Additionally, this is not a case where the Government does not have the benefit of any information
with which to form some basic methodology --the Government has already been to the pond and
fished, so to speak, with data from AOL, Yahoo, and Microsoft, and it would not have been
unreasonable at this stage to have required the Government to assist the Court in its determination of
relevance by providing the Court with more information on its plans for the information sought from
Google.

1 information *only* to perform a study, in the aggregate, of trends on the Internet" (Reply at 1:19-20)
2 (emphasis added), with no explanation of how an aggregate study of Internet trends would be
3 reasonably calculated to lead to admissible evidence in the underlying suit where the efficacy of
4 filtering software is at issue.

5 As the court in Northwestern Memorial colorfully noted, "and of course, pretrial discovery is
6 a fishing expedition and one can't know what one has caught until one fishes [b]ut Fed.R.Civ.P.
7 45(c) allows the fish to object, and when they do so the fisherman has to come up with more," 362
8 F.3d at 931 --it is difficult for a court to determine the relevance of information where the party
9 seeking the information does not concretely disclose its plans for the information sought. Given the
10 broad definition of relevance in Rule 26, and the current narrow scope of the subpoena, despite the
11 vagueness with which the Government has disclosed its study, the Court gives the Government the
12 benefit of the doubt. The Court finds that 50,000 URLs randomly selected from Google's data base
13 for use in a scientific study of the effectiveness of filters is relevant to the issues in the case of
14 ACLU v. Gonzales.⁵

15 2. Search Queries

16 In its original subpoena the Government sought a listing of the text of all search queries
17 entered by Google users over a two month period. As defined in the Government's subpoena,
18 "queries" include only the text of the search string entered by a user, and not "any additional
19 information that may be associated with such a text string that would identify the person who
20 entered the text string into the search engine, or the computer from which the text string was
21 entered." (Subpoena at 4.) The Government has narrowed its request so that it now seeks only a
22 sample of 5,000 such queries from Google's query log. The Government discloses its plans for the
23 query log information as follows: "A random sample of approximately 1,000 Google queries from a
24 _____

25 ⁵To the extent that the Government is gathering this information for some other purpose than
26 to run the sample of Google's search index through various filters to determine the efficacy of those
27 filters, the Court would take a different view of the relevance of the information. For example, the
28 Court would not find the information relevant if it is being sought just to characterize the nature of
the URL's in Google's database.

1 one-week period will be run through the Google search engine. A human being will browse the top
2 URLs returned by each search and categorize the sites by content." (Supp. Stark Decl. ¶ 3.) To the
3 extent that the URLs obtained by the researchers as a result of running the search queries provided
4 are then used to create "a sample of a relevant population of websites that can be categorized and
5 used to test filtering software" (Reply at 5) similar to the sample created from URLs from Google's
6 search index, the Court finds that were the Government to run these URLs through the filtering
7 software and analyze the results, the information sought would be reasonably calculated to lead to
8 admissible evidence.

9 Google's arguments challenging the relevance of the search queries to the Government's
10 study center around its contention that a number of additional factors exist which may mitigate the
11 correlation between a search query and the search result. (Google's Opposition to the Government's
12 Motion to Compel ("Opp."), Docket Item No. 12 at 6:9-8:1.) In particular, Google cites to the
13 presence of a safe search filter, customized searches, or advanced preferences all potentially
14 activated at the user end and not reflected in the user's search string. (Opp. at 6:17-7:2.) Google
15 also argues that the list of search queries does not distinguish between sources of the queries such as
16 adults, minors, automatic queries generated by a program, known as "bot" queries, and artificial
17 queries generated by individual users. (Opp. at 7:3-22.) Contrary to Google's belief, the broad
18 standard of relevance under Rule 26 does not require that the information sought necessarily be
19 directed at the ultimate fact in issue, only that the information sought be reasonably calculated to
20 lead to admissible evidence in the underlying litigation. See Laxalt v. McClatchy, 809 F.2d 885,
21 888 (D.C. Cir. 1987) (holding that "mere relevance to the underlying litigation" is the proper
22 standard to apply to discovery of certain FBI files). Thus, the presence of these additional factors
23 may impact the probative value of the Government's expert report in the Eastern District of
24 Pennsylvania on the effectiveness of filtering software in preventing minors from accessing
25 "harmful to minors" material on the Internet, but at this stage, the Court does not find the search
26 queries to be entirely irrelevant to the creation of a test set on which to test the effectiveness of
27

1 search filters in general.

2 **B. Undue Burden**

3 This Court is particularly concerned anytime enforcement of a subpoena imposes an
4 economic burden on a non-party. Under Rule 45(3)(a), a court may modify or quash a subpoena
5 even for relevant information if it finds that there is an undue burden on the non-party. Undue
6 burden to the non-party is evaluated under both Rule 26 and Rule 45. See Exxon Shipping Co. v.
7 U.S. Dept. of Interior., 34 F.3d 774, 779 (9th Cir. 1994).

8 1. Technological Burden of Production

9 Google argues that it faces an undue burden because it does not maintain search query or
10 URL information in the ordinary course of business in the format requested by the Government.
11 (Opp. at 16:22-15.) As a general rule, non-parties are not required to create documents that do not
12 exist, simply for the purposes of discovery. Insituform Tech., Inc. v. Cat Contracting, Inc., 168
13 F.R.D. 630, 633 (N.D. Ill. 1996). In this case, however, Google has not represented that it is unable
14 to extract the information requested from its existing systems. Google contends that it must create
15 new code to format and extract query and URL data from many computer banks, in total requiring
16 up to eight full time days of engineering time. Because the Government has agreed to compensate
17 Google for the reasonable costs of production, and given the extremely scaled-down scope of the
18 subpoena as modified, the Court does not find that the technical burden of production excuses
19 Google from complying with the subpoena. Later in this Order, the Court addresses other concerns
20 with respect to this information, however.

21 Google also argues that even if the Government compensates Google for its engineering
22 time, if the Government plans on executing a high volume of searches on Google, such searches
23 would lead to an interference with Google's search engine and disrupt use by users and advertisers.
24 (Opp. at 16:24-17:3.) The Government only intends to run 1,000 to 5,000 of the search queries
25 through the Google search engine. (Supp. Stark Decl. ¶ 4.) Furthermore, these searches will be run
26 by humans who will then categorize the search results and record their findings. (Supp. Stark Decl.

27
28

1 ¶ 4.) Given the volume and rate of the proposed study, the Court finds that the additional burden on
2 Google's search engine caused by the Government's study as represented to the Court, is likely to be
3 de minimus.

4 2. Potential for Loss of User Trust

5 Google also argues that it will be unduly burdened by loss of user trust if forced to produce
6 its users' queries to the Government. Google claims that its success is attributed in large part to the
7 volume of its users and these users may be attracted to its search engine because of the privacy and
8 anonymity of the service. According to Google, even a perception that Google is acquiescing to the
9 Government's demands to release its query log would harm Google's business by deterring some
10 searches by some users. (Opp. at 18.)

11 Google's own privacy statement indicates that Google users could not reasonably expect
12 Google to guard the query log from disclosure to the Government. Google's privacy statement at
13 www.google.com/privacypolicy.html states only that Google will protect "personal information" of
14 users. "Personal information" is expressly defined for users at www.google.com/privacy_faq.html
15 as "information that you provide to us which personally identifies you, such as your name, email
16 address or billing information, or other data which can be reasonably linked to such information by
17 Google." (Second Decl. of Joel McElvain, Ex. C.) Google's privacy policy does not represent to
18 users that it keeps confidential any information other than "personal information." Neither Google's
19 URLs nor the text of search strings with "personal information" redacted, are reasonably "personal
20 information" under Google's stated privacy policy. Google's privacy policy indicates that it has not
21 suggested to its users that non-"personal information" such as that sought by the Government is kept
22 confidential.

23 However, even if an expectation by Google users that Google would prevent disclosure to
24 the Government of its users' search queries is not entirely reasonable, the statistic cited by Dr. Stark
25 that over a quarter of all Internet searches are for pornography (Supp. Stark Decl. ¶4), indicates that
26
27

1 at least some of Google's users expect some sort of privacy in their searches.⁶ The expectation of
 2 privacy by some Google users may not be reasonable, but may nonetheless have an appreciable
 3 impact on the way in which Google is perceived, and consequently the frequency with which users
 4 use Google. Such an expectation does not rise to the level of an absolute privilege, but does indicate
 5 that there is a potential burden as to Google's loss of goodwill if Google is forced to disclose search
 6 queries to the Government.

7 3. Trade Secret

8 Rule 45(c)(3)(B) provides additional protections where a subpoena seeks trade secret or
 9 confidential commercial information from a nonparty. Once the nonparty shows that the requested
 10 information is a trade secret or confidential commercial information, the burden shifts to the
 11 requesting party to show a "substantial need for the testimony or material that cannot be otherwise
 12 met without undue hardship and assures that the person to whom the subpoena is addressed will be
 13 reasonably compensated." Rule 45(c)(3)(B). Upon such a showing, "the court may order
 14 appearance or production only upon specified conditions." Id. See also Klay v. Humana, 425 F.3d
 15 977, 983 (11th Cir. 2005); Heat & Control, Inc. v. Hester Industries, Inc., 785 F.2d 1017, 1025 (Fed.
 16 Cir. 1986).

17 a. Search Index and Query Log as Trade Secrets

18 Trade secret or commercially sensitive information must be "important proprietary
 19 information" and the party challenging the subpoena must make "a strong showing that it has
 20 historically sought to maintain the confidentiality of this information." Compaq Computer Corp. v.
 21 Packard Bell Elec., Inc., 163 F.R.D. 329, 338 (N.D. Cal. 1995). A statistically significant sample of
 22

23
 24 ⁶ At the hearing, the Government argued that Google should not be concerned about loss of
 25 user trust because Google already discloses its users' search queries on Google Zeitgeist. Had the
 26 Government truly believed that substantial amounts of search query information could be obtained
 27 from Google Zeitgeist, it is unlikely that the Government would require further search query
 information from Google. On the Court's examination of Google Zeitgeist at
<http://www.google.com/press/zeitgeist.html>, the website only provides the top ten search queries by
 country or the top fifteen gaining search queries in the United States. These queries for the Week of
 March 13, 2006, include "teri hatcher," "world baseball classic," and "sopranos."

1 Google's search index and Google's query log would have independent economic value from not
2 being known generally to the public. The disclosure of a statistically significant sample of Google's
3 search index or query log may permit competitors to estimate information about Google's indexing
4 methods or Google's users. (Decl. of Matt Cutts ("Cutts Decl.") ¶¶ 26, 27.) By declaration, Google
5 represents that it does not share this information with third parties and it has security procedures to
6 maintain the confidentiality of this information. (Cutts Decl. ¶¶ 29-35; Decl. of Marty Lev.)

7 At oral argument, counsel for Google acknowledged that samples from its proprietary search
8 index and query log of 50,000 URLs and 5,000 search queries are far less likely to lead to trade
9 secret disclosure than the Government's original requests. Because Google still continues to claim
10 information about its entire search index and entire query log as confidential, the Court will presume
11 that the requested information, as a small sample of proprietary information, may be somewhat
12 commercially sensitive, albeit not independently commercially sensitive. Successive disclosures,
13 whether in this lawsuit or pursuant to subsequent civil subpoenas, in the aggregate could yield
14 confidential commercial information about Google's search index or query log.

15 b. Entanglement in the Underlying Litigation

16 Google's remaining trade secret argument is that despite the narrowness of the sample
17 provided, it would become entangled in the underlying litigation where further discovery would risk
18 trade secret disclosure. Rule 45(c)(3)(B) was intended to provide protection for the intellectual
19 property of nonparties. See Mattel, Inc. v. Walking Mountain Prod., 353 F.3d 792, 814 (9th Cir.
20 2003) (citing Rule 45 advisory committee's notes (1991)). On the one hand, a determination of the
21 propriety of further discovery is for another set of motions, and not the one presently before the
22 Court. On the other hand, further discovery in this case that would require disclosure of Google's
23 trade secrets is not merely a remote possibility. The Government has represented that it has
24 sufficient information from other search engines with which to perform its study, but seeks
25 information from Google because such information would add "substantial luster" to its study --
26 ostensibly because there is something unique about the world of Google. The nature and extent of

27

28

1 that uniqueness, if sufficient to add substantial luster to the Government's study, is also likely to be a
2 matter of discovery for Plaintiffs in the underlying suit involving more than the Government's
3 proposed "fifteen-minute deposition" of a Google engineer to confirm that the statistician's
4 procedure had been followed.

5 In light of the comments of Plaintiffs' counsel at the hearing, the Court can foresee further
6 entanglement based on Plaintiffs' challenge to the Government's ultimate study. In litigation where
7 the ultimate question is not whether there is adult material on the Internet, but fundamentally about
8 limiting the access by minors to such adult material, it is quite likely that Plaintiffs will challenge the
9 sample produced by Google as not representative of what minors search for or encounter on the
10 Internet. Such an inquiry would require additional discovery, some of which may implicate
11 Google's confidential commercial information. At the hearing, Plaintiffs' counsel stated that it had
12 already commenced such discovery with respect to a search engine included in the Government's
13 study. In other words, this Court is concerned that a narrow sample of Google's proprietary index
14 and query log, while in itself not likely to lead to the disclosure of confidential information, may act
15 as the thin blade of the wedge in exposing Google to potential disclosure of its confidential
16 commercial information.

17 c. Substantial Need

18 The burden thus shifts to the Government to demonstrate that the requested discovery is
19 relevant and essential to a judicial determination of its case. See Upjohn Co. v. Hygieia Biological
20 Laboratories, 151 F.R.D. 355, 358 (E.D. Cal. 1993). Because "there is no absolute privilege for
21 trade secrets and similar confidential information," Centurion Indus., Inc. v. Warren Steurer and
22 Assoc., 665 F.2d 323, 325 (10th Cir. 1981) (citing Federal Open Market Committee v. Merrill, 443
23 U.S. 340, 362 (1979)), the district court's role in this inquiry is to balance the need for the trade
24 secrets against the claim of injury resulting from disclosure. Heat & Control, 785 F.2d at 1025. The
25 determination of substantial need is particularly important in the context of enforcing a subpoena
26 when discovery of trade secret or confidential commercial information is sought from non-parties.

27

28

1 See Mattel, 353 F.3d at 814.

2 Google contends that it should not be compelled to produce its search index or query log
3 because the information sought by the Government is readily available from open URL databases
4 such as Alexa and transparent search engines such as Dogpile, or that the Government already has
5 sufficient information from AOL, Yahoo, and Microsoft. As a rule, information need not be
6 dispositive of the entire issue disputed in the litigation in order to be discoverable by subpoena. See
7 Compaq, 163 F.R.D. at 333 n.25. In Compaq, industry practice was a material issue in the lawsuit,
8 and the court refused to quash a subpoena for information from a non-party industry member based
9 on the non-party's argument that information could be discoverable from other industry members.
10 Id. Similarly, at oral argument, the Government's counsel likened its discovery goals to a team of
11 researchers studying an elephant by separately viewing the trunk, the ears, the tail, etc., and piecing
12 the research together to get a picture of the elephant as whole.

13 In this case, the Government has demonstrated a substantial need for some information from
14 Google in creating a set of URLs to run through filtering software. It is uncontested that Google is
15 the market leader with over 45% of the search engine market. (Supp. Stark Decl. ¶¶ 4-5.) Because
16 Google has the greatest market share, the Government's study may be significantly hampered if it
17 did not have access to some information from the most often used search engine.

18 4. Cumulative and Duplicative Discovery

19 What the Government has not demonstrated, however, is a substantial need for *both* the
20 information contained in the sample of URLs and sample of search query text. Furthermore, even if
21 the information requested is not a trade secret, a district court may in its discretion limit discovery
22 on a finding that "the discovery sought is unreasonably cumulative or duplicative, or is obtainable
23 from some other source that is more convenient, less burdensome, or less expensive." Rule
24 26(b)(2)(i). See In re Sealed Case (Medical Records), 381 F.3d 1205, 1215 (D.C. Cir. 2004) (citing
25 the advisory committee's notes to Rule 26 and finding that "the last sentence of Rule 26(b)(1) was
26 added in 2000 'to emphasize the need for active judicial use of subdivision (b)(2) to control

27

28

1 excessive discovery"). From this Court's interpretation of the Government's general statements of
2 purpose for the information requested, both the sample of URLs and the set of search queries are
3 aimed at providing a list of URLs which will be categorized and run through the filtering software in
4 an effort to determine the effectiveness of filtering software as to certain categories. Both sources of
5 the URL "test set" list seem to be open to the same sorts of criticism by Plaintiffs in the underlying
6 litigation. The content of these objections are not germane to the Court's determination of whether
7 the information sought is relevant under the broad dictates of Rule 26, but the actual similarity of the
8 two categories of information sought in their presumed utility to the Government's study indicates
9 that it would be unreasonably cumulative and duplicative to compel Google to hand over both sets
10 of proprietary information. To borrow the Government's vivid analogy, in order to aid the
11 Government in its study of the entire elephant, the Court may burden a non-party to require
12 production of a picture of the elephant's tail, but it is within this Court's discretion to not require a
13 non-party to produce another picture of the same tail.

14 Faced with duplicative discovery, and with the Government not expressing a preference as to
15 which source of the test set of URLs it prefers, this Court exercises its discretion pursuant to Rule
16 26(b)(2) and determines that the marginal burden of loss of trust by Google's users based on
17 Google's disclosure of its users' search queries to the Government outweighs the duplicative
18 disclosure's likely benefit to the Government's study. Accordingly, the Court grants the
19 Government's motion to compel only as to the sample of 50,000 URLs from Google's search index.

20 **C. Protective Order**

21 As trade secret or confidential business information, Google's production of a list of URLs to
22 the Government shall be protected by protective order. Generally, "the selective disclosure of
23 protectable trade secrets is not per se 'unreasonable and oppressive,' when appropriate protective
24 measures are imposed." Heat & Control, 785 F.2d at 1025. The Court recognizes that Google was
25 unable to negotiate the particular provisions of the protective order in the underlying litigation,
26 (Opp. at 12:15-18) but since Google's filing of its Opposition, the Government has considerably
27
28

1 narrowed its request for Google's information from its proprietary search index such that the risk of
2 trade secret disclosure is substantially mitigated.

3 The Court grants the motion to compel as to a set of 50,000 URLs from Google's search
4 index and orders the parties to show cause, if any, on or before April 3, 2006, why a designation of
5 the produced information as "Confidential" under the existing protective order is insufficient
6 protection for Google's confidential commercial information.

7 **D. Privacy**

8 The Court raises, sua sponte, its concerns about the privacy of Google's users apart from
9 Google's business goodwill argument. In Gill v. Gulfstream Park Racing Assoc., the First Circuit
10 held that "considerations of the public interest, the need for confidentiality, and privacy interests are
11 relevant factors to be balanced" in a Rule 26(c) determination regarding the subpoena of documents
12 used to prepare an allegedly defamatory report issued by a non-party trade association. 399 F.3d
13 391, 402 (1st Cir. 2005) (citing, as also concerned with the interest of privacy in the context of
14 discovery, Seattle Times Co. v. Rhinehart, 467 U.S. 20, 35 n.21 (1984), In re Sealed Case (Medical
15 Records), 381 F.3d at 1215, and Ellison v. Am. Nat'l Red Cross, 151 F.R.D. 8, 11 (D.N.H. 1993)).

16 The Government contends that there are no privacy issues raised by its request for the text of
17 search queries because the mere text of the queries would not yield identifiable information.
18 Although the Government has only requested the text strings entered (Subpoena at 4), basic
19 identifiable information may be found in the text strings when users search for personal information
20 such as their social security numbers or credit card numbers through Google in order to determine
21 whether such information is available on the Internet. (Cutts Decl. ¶¶ 24-25.) The Court is also
22 aware of so-called "vanity searches," where a user queries his or her own name perhaps with other
23 information. Google's capacity to handle long complex search strings may prompt users to engage
24 in such searches on Google. (Cutts Decl. ¶ 25.) Thus, while a user's search query reading "[user
25 name] stanford glee club" may not raise serious privacy concerns, a user's search for "[user name]
26 third trimester abortion san jose," may raise certain privacy issues as of yet unaddressed by the
27

1 parties' papers. This concern, combined with the prevalence of Internet searches for sexually
2 explicit material (Supp. Stark Decl. ¶ 4) --generally not information that anyone wishes to reveal
3 publicly --gives this Court pause as to whether the search queries themselves may constitute
4 potentially sensitive information.

5 The Court also recognizes that there may a difference between a private litigant receiving
6 potentially sensitive information and having this information be produced to the Government
7 pursuant to civil subpoena. The interpretation of the Federal Rules in this Circuit requires that
8 "when the government is named as a party to an action, it is placed in the same position as a private
9 litigant, and the rules of discovery in the Federal Rules of Civil Procedure apply." Exxon Shipping,
10 34 F.3d at 776 n.4. However, in Exxon Shipping, the Ninth Circuit was faced with a situation where
11 a litigant sought discovery *from* the Government; in this case, information is being produced *to* the
12 Government. Even though counsel for the Government assured the Court that the information
13 received will only be used for the present litigation, it is conceivable that the Government may have
14 an obligation to pursue information received for unrelated litigation purposes under certain
15 circumstances regardless of the restrictiveness of a protective order.⁷ The Court expressed this
16 concern at oral argument as to queries such as "bomb placement white house," but queries such as
17 "communist berkeley parade route protest war" may also raise similar concerns. In the end, the
18 Court need not express an opinion on this issue because the Government's motion is granted only as
19 to the sample of URLs and not as to the log of search queries.

20 **E. Electronic Communications Privacy Act**

21 The Court also refrains from expressing an opinion on the applicability of the Electronic
22 Communications Privacy Act, codified at 18 U.S.C. §§ 2510 to 2712. The ECPA was enacted in
23 1986 "to update and clarify federal privacy protections and standards in light of dramatic changes in
24 new computer and telecommunication technologies." Freedman v. America Online, Inc., 303 F.

25
26 ⁷ "Says the DOJ's [spokesperson Charles] Miller, "I'm assuming that if something raised
27 alarms, we would hand it over to the proper [authorities]." (Decl. of Ashok Ramani, Ex. B.
"Technology: Searching for Searches," Newsweek, Jan. 30, 2006.) (second alteration in original)

1 Supp. 2d 121, 124 (D. Conn. 2004) (quoting 132 CONG. REC. S. 14441 (1986)). See also Theofel v.
2 Fare-Jones, 359 F.3d 1066, 1071 (9th Cir. 2004). The Court only notes that the ECPA does not bar
3 the Government's request for sample of 50,000 URLs from Google's index though civil subpoena.

4 V. CONCLUSION

5 As expressed in this Order, the Court's concerns with certain aspects of the Government's
6 subpoena have been mitigated by the reduced scope the Government's present requests. Nothing in
7 this Order is intended to indicate how the Court would rule on the original broad subpoena or on any
8 follow-up subpoena. The Court's decision on this Motion to Compel reflects the limited use to
9 which the Government intends to put the information produced in response to the subpoena. In
10 particular, this Order does not address the Plaintiffs' concern articulated at the hearing about the
11 appropriateness of the Government's use of the Court's subpoena power to gather and collect
12 information about what individuals search for over the Internet.

13 With these limitations, for the reasons stated in this Order, unless the parties agree otherwise
14 on or before April 3, 2006, Google is ordered to confer with the Government to develop a protocol
15 for the random selection and afterward immediate production of a listing of 50,000 URLs in
16 Google's database on the following conditions:

- 17 1. In the development or implementation of the protocol, Google shall not be required to
18 disclose proprietary information with respect to its database;
- 19 2. The Government shall pay the reasonable cost incurred by Google in the formulation and
20 implementation of the extraction protocol;
- 21 3. Any information disclosed in response to this Order shall be subject to the protective
22 order in the underlying case;

23 To the extent the motion seeks an order compelling Google to disclose search queries of its
24 users the motion is DENIED. The Court retains jurisdiction to enforce this Order.

25 Dated: March 17, 2006

26 06cv80006subpoena

27 /s/ James Ware
28 JAMES WARE
United States District Judge

1 **THIS IS TO CERTIFY THAT COPIES OF THIS ORDER HAVE BEEN DELIVERED TO:**

2 Joel McElvain: joel.l.mcelvain@usdoj.gov
3 Albert Gidari, Jr.: agidari@perkinscoie.com
4 Lisa Delehunt: ldelehunt@perkinscoie.com

5 **Dated: March __, 2006**

Richard W. Wieking, Clerk

6 **By: /s/ JW Chambers _____**
7 **Melissa Peralta**
8 **Courtroom Deputy**