

1 PETER D. KEISLER
Assistant Attorney General
2 THEODORE HIRT
Assistant Branch Director
3 JOEL McELVAIN, D.C. Bar No. 448431
Trial Attorney
4 U.S. Department of Justice
Civil Division, Federal Programs Branch
5 20 Massachusetts Ave., NW
Washington, DC 20001
6 Telephone: (202) 514-2988
7 Fax: (202) 616-8202
Email: Joel.L.McElvain@usdoj.gov

8 Attorneys for Alberto R. Gonzales

9
10 IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
(SAN JOSE DIVISION)

11 **ALBERTO R. GONZALES, in his official)**
12 **capacity as ATTORNEY GENERAL OF THE)**
13 **UNITED STATES,**

14 **Movant,**

15 v.

16 **GOOGLE INC.,**

17 **Respondent.**

Case No. _____-MISC

**Notice of Filing of Declaration of
Philip B. Stark, Ph.D.**

18 NOTICE is hereby given of the filing by Alberto R. Gonzales, in his official
19 capacity as Attorney General of the United States, of the attached Declaration of Philip B.
20 Stark, Ph.D. This declaration is filed in support of the Motion to Compel Compliance
21

1 with Subpoena Duces Tecum, in which the Attorney General seeks to compel Google Inc.
2 to comply with the subpoena issued to it pursuant to Federal Rule of Civil Procedure 45.

3 Dated: January 18, 2006

Respectfully submitted,

4 PETER D. KEISLER
5 Assistant Attorney General

6 THEODORE HIRT
Assistant Branch Director

7 

8 JOEL McELVAIN
9 Trial Attorney
10 United States Department of Justice
11 Civil Division, Federal Programs Branch
12 20 Massachusetts Ave., NW, Room 7130
13 Washington, D.C. 20001
14 Telephone: (202) 514-2988
15 Fax: (202) 616-8202
16 Email: Joel.L.McElvain@usdoj.gov

17 *Attorneys for the Movant, Alberto R. Gonzales*

DECLARATION OF PHILIP B. STARK, PH.D.

1. I am Professor of Statistics at the University of California at Berkeley. I have been asked by the U.S. Department of Justice, in connection with the case *ACLU v. Gonzales*, no. 98-5591 (E.D. Pa.), to devise studies of sites on the world wide web (URLs) available through search engines and of users' queries to search engines, and to analyze the resulting data.

2. Together with attorneys from the U.S. Department of Justice, I have been involved in conversations with attorneys from Google and in conversations with attorneys and engineers from other search providers to develop practical approaches to sampling their databases of URLs and user queries. I have started to analyze the samples produced by search providers other than Google.

3. Reviewing URLs available through search engines will help us understand what sites users can find using search engines, to estimate the prevalence of harmful-to-minors (HTM) materials among such sites, to characterize those sites, and to measure the effectiveness of content filters in screening HTM materials from those sites.

4. Reviewing user queries to search engines will help us understand the search

behavior of current web users, to estimate how often web users encounter HTML materials through searches, and to measure the effectiveness of filters in screening those materials.

5. According to a recent study, Google has the largest share of the web search market.¹ We seek to understand and characterize URLs available through search engines and users' queries to search engines. Google specifically—and search engines more generally—can give the best possible information about what is available through search engines, and about what users search for.

6. A sample of the URLs in Google's database would increase our coverage of the part of the web available through search engines, making our sample of URLs more comprehensive and providing a check against the samples provided by other vendors.

7. Because it is plausible that users of different search engines have different browsing habits, a sample of queries run on Google's search engine combined with samples from other search engines will help us ensure that our overall sample of queries is representative of the public at large.

¹ See comScore, <http://www.comscore.com/press/release.asp?press=622>

8. Google objects that producing samples of URLs and of user queries poses a burden. Other search providers have produced electronic files containing random samples of 1 million URLs and a full week of user queries, and did not complain of undue burden.

It might be that Google's architecture makes it more difficult or costly to sample their database of URLs or to sample their user queries than it was for the other search providers; if so, I can work with them to design a multi-stage sample that would reduce any burden below that incurred by the other search providers.

9. Google objects that producing a sample of queries would compromise user privacy. Other vendors have been able to produce samples of queries with all information that might identify a user removed.

10. Google claims that producing a sample of URLs would create a false impression that their database is representative of the entire world-wide web. But we do not seek to study the entire world-wide web: our goal is to characterize sites that can be found using search engines, and what people see in practice when they use search engines. Google is one of the most popular search engines. Therefore, what Google has in its database of URLs, and what searches users run on Google, are directly relevant.

I declare under penalty of perjury that the foregoing is true and correct.

 _____

Dated 16 January 2006.

Philip B. Stark

