UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF FLORIDA

CASE NO. 11-20427-WILLIAMS/TURNOFF

DISNEY ENTERPRISES, INC.,
TWENTIETH CENTURY FOX FILM CORPORATION,
UNIVERSAL CITY STUDIOS PRODUCTIONS LLLP,
COLUMBIA PICTURES INDUSTRIES, INC., and
WARNER BROS. ENTERTAINMENT INC.,

Plaintiffs,

v.

HOTFILE CORP., ANTON TITOV, and
DOES 1-10.

  Defendants.

_____/

HOTFILE CORP.,

Counterclaimant,

v.

WARNER BROS. ENTERTAINMENT INC.,

Counterdefendant.

_____/

DECLARATION OF DR. IAN FOSTER IN SUPPORT OF WARNER'S MOTION FOR SUMMARY JUDGMENT

PUBLIC REDACTED VERSION

1.   My name is Ian Foster and I currently hold the position of Director of the
Computation Institute at Argonne National Laboratory and the University of Chicago. I also
hold the positions of Arthur Holly Compton Distinguished Service Professor of Computer
Science at the University of Chicago, and of Argonne Distinguished Fellow at the Argonne

National Laboratory. My professional and academic background include extensive experience in designing, operating, maintaining, and studying large, distributed computer systems, including the management of large databases required for such systems to operate. A copy of my *curriculum vitae*, including lists of my previous publications, is attached hereto as Exhibit A. I have not given testimony in any legal proceeding the past four years.

2.   I have been asked by Warner, for purposes of Hotfile's counterclaim, to extract and describe certain data from data sets maintained and produced by www.hotfile.com ("Hotfile"). The observations and conclusions set forth below are based on my own observation and use of Hotfile's produced data, as informed by my specialized knowledge, education, and expertise as applied to the facts and circumstances in this case. In addition, I have for purposes of this declaration consulted and reviewed the following data sets and documents produced by Hotfile (including later-supplemented versions):

     a. Actiondat.csv;

     b. Affpay.csv;

     c. Uploads.csv;

     d. Userdat.csv;

     e. Dmcanotices.csv;

     f. Users_cowner_upload.csv;

     g. Strikes.csv;

     h. File.csv;

     i. Uploadsurl.csv;

     j. HF0285583;

     k. Takedown notices sent to Hotfile produced by Hotfile in this litigation;

     l. Transcripts and selected exhibits from of the November 16, 2011 and December 5-8 depositions of Anton Titov;

     m. Source code related to the removal and blocking of files produced by Hotfile.

3.   I am being compensated for my study and testimony in this case at a rate of $500 per hour. If called as a witness at trial, I would testify as to the contents of this declaration.

Methodology

4.   Hotfile provided various data sets in this case, which I understand were extracted from its database(s), in comma separated value (or "CSV") format, a common format used to

represent tabular data. Analyzing and querying data sets in this format is a straightforward proposition: they were loaded to a database server (using MySQL, a common database management system); I then constructed and indexed tables (which associate subsets of the data based on defined criteria) using the MySQL software, and then ran queries using common database commands as discussed below.

Number of Takedown Notices Received by Hotfile

5.   Included in Hotfile's data production were various data sets and other sources of information reflecting takedown notices Hotfile has received from copyright owners. I have been asked by Warner's counsel to provide the number of such notices received. Hotfile has provided data reflecting notices referring to a cumulative total of approximately ten million files (and more than a million after the filing of this lawsuit on February 8, 2011). The methodology for calculating this number is as follows.

6.   First, Hotfile provided a data set "dmcanotices.csv" (as well as subsequent supplements) containing entries corresponding to takedown notices that Hotfile has received by email. These can be summed using common database commands. Second, Hotfile produced another data set – "users_cowner_upload.csv" – containing records of takedown notices sent by means of Hotfile's web-based interface or special rightsholder account ("SRA"), which can likewise be summed using common database commands. In theory, these two sources ought to be sufficient to capture takedown notices received by Hotfile. In analyzing the dmcanotices.csv data set, however, I observed that the data set only contains takedown notices from July 2, 2009 through September 15, 2009, and then again from September 19, 2010 through the present. In other words, a year of data (running from September 2009 through September 2010) is missing from the dmcanotices.csv data set entirely. I understand that Hotfile has stated that this data was lost and cannot be recovered.

7.   Although this data is missing from the dmcanotices.csv data set, it can be substantially reconstructed from the actual email takedown notices that Hotfile received and produced, many of which were provided in text form. Using a script (i.e., a small computer program), I extracted the "uploadid" (i.e., the unique numerical identifier that Hotfile assigns to each file uploaded to its service) of the files contained in the accessible takedown notices for the

period for which Hotfile's dmcanotices.csv data set was missing entries entirely, then summed those and added them to the total count from the other two sources.[1]

Number of Users Suspended by Hotfile for Repeat Infringement Subsequent to Litigation

8.   I have been asked by Warner's counsel to provide a count, from Hotfile's data, of the total number of users Hotfile suspended based on alleged copyright infringement subsequent to the filing of Plaintiffs' lawsuit against Hotfile.

9.   Hotfile maintains records of its suspension of users based on alleged copyright infringement. One of Hotfile's produced data sets, "actiondat.csv" (and supplements to that data set) contains records of actions Hotfile took regarding users. The actiondat data set associates the unique identification number that Hotfile assigns to each user ("userid") with the "type" of action (which includes "suspenduser"), the date of the action ("dtaction") and other information about the suspension, including the reason for the suspension ("params"). Therefore, this data set readily identifies users whom Hotfile terminated based on alleged copyright infringement subsequent to the filing of the Complaint on February 8, 2011. Such suspensions – which began on February 18, 2011 – exceed 22,000 as of the latest supplement to Hotfile's actiondat data set.

Data Regarding Counterclaim Files and Their Uploading Users

10.   I understand that Plaintiffs' expert Scott Zebrak is submitting a declaration containing information concerning each of the 890 files that Hotfile is claiming that Warner wrongfully took down. At Warner's counsel's request, I have extracted and provided certain data from Hotfile's produced data sets, for use in connection with Mr. Zebrak's report, that relate to each of the files identified by Hotfile in the Counterclaim. This includes, for each file:

      a. The unique identifier ("userid") that Hotfile assigns to the uploading user, as well as the name of the user's account. Both values are straightforward associations of data contained in Hotfile's userdat and uploads data sets (which contain various data concerning users and concerning uploaded files, respectively).

      b. Whether or not a file identical to the file identified in the Counterclaim (i.e., a file with the same "hash" value)[2] had been the subject of a takedown notice sent by a

---

[1] I note here that my methodology was conservative in that it was limited to takedown notices accessible in text form. Hotfile produced a much smaller number of notices in other formats that I did not consider due to the additional difficulty and higher potential for introducing inadvertent errors into the count. Therefore, the actual number of takedown notices received by Hotfile is likely somewhat higher.

copyright owner other than Warner. This can be determined by identifying the hash values of the files in the Counterclaim (which are represented as "md5"and "sha1" in the uploads and file data sets), then comparing those hash values to the hash values of files on which Hotfile received a takedown notice (excluding Warner's own takedowns through the SRA), and identifying any matches.[3]

c. Search results reflecting the names of files with similar names that had been removed from Hotfile for reasons related to claimed infringement.

    i. This can be determined by creating a table of the files removed from Hotfile due to infringement claims. This includes files on which Hotfile received a takedown notice (the process for identifying which I described in Paragraphs 5-7 above) as well as files that Hotfile itself deleted (which includes files removed in response to a specific takedown notice, but also includes, e.g., files Hotfile removed due to takedown notices that removed multiple files at once, files Hotfile removed or blocked due to a hash match with a different file removed due to claimed infringement, etc.).

    ii. This latter category can be shown by the file's "status" in Hotfile's uploads data set; Hotfile (per the testimony of Anton Titov and associated deposition exhibits) uses various different codes to refer to different file statuses, with status "5" and "2" reflecting Hotfile's own suspension and deletion of files (as opposed to files deleted for other specific reasons, e.g., deleted by the user or deleted due to inactivity).

    iii. Once this table was constructed, I excluded files to avoid counting files as that had been removed for reasons other than claimed infringement.

        1. First, I excluded Warner's own takedowns through the SRA.

        2. Second, as a conservative measure, unless there was a takedown notice for the specific file, I also excluded files where the user had been suspended, for any reason. This restriction is very conservative; files removed for infringement were many cases uploaded by users who were later suspended. Nevertheless, I understand from Mr. Titov's testimony, as well as from my analysis of the data, that Hotfile appears to have for certain time periods and in certain cases used statuses "5"and "2" to designate files suspended or deleted because the user was suspended, even in the absence of alleged infringement with respect to the specific file itself. Therefore, to err on the side of underinclusion rather than overinclusion, I did not include such files in the table.

        3. Third, I excluded files where Hotfile had chosen not to "block" the file from future upload (i.e., where the "blocked" value for the file

---

[2] A "hash" does not purport to recognize when two different files contain the same content (since the representation of that content in a digital file can differ). However, it does usefully identify when two digital files are exact copies of one another.

[3] MD5 and SHA1 are both common algorithms for identifying a file's hash used by Hotfile.

in Hotfile's uploads data set was set to a value of "0"). Again, this is a conservative measure, as Hotfile may have declined to block uploads of files removed for reasons related to infringement in some cases. Nevertheless, to err on the side of underinclusion rather than overinclusion, I did not include files where Hotfile had itself chosen not to block future uploads of the file.

iv. Once this table was constructed, the identification of file name matches is a simple matter of using MySQL's search functionality to identify candidate matches. Here, I used search terms for the 890 counterclaim files and ran them against the files in the table constructed and narrowed as described above. I then provided up to 25 candidate matches for each file to Plaintiffs for analysis.

d. Search results reflecting the names of copyright owners, with names similar to the copyright owners that Plaintiffs had identified as being the owners of the files in question, that had sent takedown notices to Hotfile. I conducted this search by using the search functionality of MySQL to search both the bodies of the takedown notices reflected in the dmcanotices data set and the usernames and email addresses of copyright owners in the users_cowner_upload data set. I then provided up to 25 candidate matches for each file to Plaintiffs for analysis.

11. In addition, also from Hotfile's data sets, I have extracted at counsel's request the following data concerning the uploading user of each file:

a. How many copyright "strikes" Hotfile had itself assigned to the user based on its own count of takedown notices sent against the user's files (a straightforward association of the userid with Hotfile's "strikes" data set).

b. The number of discrete days on which Hotfile received a takedown notice for a file uploaded by the user (which I have termed "Notice Days"). This can be calculated by associating takedown notices with the user, then in turn summing the distinct number of days on which such notices are dated.

c. Whether or not the user was suspended for copyright infringement, and if so, the date of the first such suspension. As described in paragraph 9 above, Hotfile maintains records of suspension events, dates, and reasons in its actiondat data set. It also maintains records (in the userdat data set) of whether or not a user is suspended. In some cases, Hotfile's userdat data reflects the suspension of a user but the actiondat data set did not contain an entry reflecting the reason. In those cases, I have indicated "Reason Unknown" ("RU").

d. The number of days, if any, that elapsed between the user's suspension and the date of the Warner notice at issue (calculated by means of a simple script reflecting the difference between the date of the notice and the date of the suspension).

e. The number of discrete days after February 18, 2011 on which the takedown notices over which Hotfile is suing in its counterclaim were sent against a file

uploaded by the user. (This is a simple sorting and/or manual calculation within Microsoft Excel).

f. Whether or not the user is a "Premium" user, and, if so, until when. This data is reflected in the userdat data set.

g. Whether Hotfile has paid the user as an "Affiliate" and, if so, the sum of such affiliate payments. This data is reflected in Hotfile's userdat and affpay data sets.

h. Whether Hotfile has a copy of the file. Hotfile maintains data in the uploads data set showing the size of each file; whether or not Hotfile has such data can act as a proxy for whether Hotfile has a copy of the file itself.

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

Executed in the State of Illinois this 10th day of February, 2012.

_____
Ian Foster, Ph.D.