Solutions to Homework 4

Question 1.

(a) (7 Points) "Peer production" is production by users who are peers (*i.e.*, none is above another in a formal organizational hierarchy) and who self-identify and self-organize by communicating directly with each other. Benkler distinguishes it from two other well-studied ways in which production occurs, namely markets (in which communication is facilitated by prices) and hierarchies (in which communication flows along a chain of command). He claims that peer production is superior to other production modes for Internet-scale information products, because "there are increasing returns to scale of the set of agents permitted to work with a set of resources in pursuit of projects, and to the set of resources agents are allowed to work with. The unbounded sets of both human capital and information inputs that can be used in peer production capture these economies of scale more effectively than can firms, and to a lesser extent markets, both of which rely on securing access to bounded sets of agents and information inputs to reduce uncertainty about the success of projects."

(b) (6 Points) Examples include Linux, BSD Unix, Apache, GNU Emacs, Perl, and Sendmail.

(c) (4 Points) iii. trespass

(d) (4 Points) Feist Publications, Inc. *v*. Rural Telephone Services Co., Inc.

(e) (4 Points) i. low-cost communication and low-cost access to information

Question 2.

(a) (5 Points) One basic reason is that spiders use computational resources that might be needed by the organization that is hosting the web page. Other plausible reasons will also be given full credit.

(b) (5 Points) As explained in "How Internet Search Engines Work," careless or unscrupulous page owners might add meta-tags that match very commonly used search terms but that have nothing to do with the contents of the pages.

(c) (3 Points) Examples include "gopher," "Archie," and "Veronica."

(d) (8 Points) As explained in the *Google Press Reviewer's Guide*, the "PageRank" algorithm "relies on the democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from Page A to Page B as a vote by Page A for Page B. And Google looks at more than the sheer volume of votes or links a page receives; it also analyzes the pages that cast the votes. Votes cast by pages that are themselves important weigh more heavily and help to make other pages important."

(e) (5 Points) Both are commercial products hosted by Google; the market for them consists of organizations that need search capabilities. An owner of a portal or website can use Google WebSearch™ to offer *web-wide* search capability to its users. It can use Google SiteSearch™ to enable search *within its specific site.*

Question 3.

(a) (5 Points) The "abundance problem" occurs when search-engine users submit "broad-topic" queries (*e.g.*, "cars," "computers," or "Afghanistan"); the crux of it is that the number of web pages that could reasonably be considered relevant is far too large for a human user to digest. Its opposite is the "scarcity problem." This occurs when a user submits a very specific query (*e.g.*, "Netscape support for the JDK 1.1 code-signing API"); there are very few relevant web pages, and it is difficult for search engines to find them. (The scarcity problem is not part of the homework question, and no definition of it is needed for full credit. The definition is included here just FYI ["for your information"].)

(b) (10 Points) A "hub" is a web page that contains good links, and an "authority" is a web page that contains good content. Hub-authority relationships are "mutually reinforcing" in that the higher the hub score (respectively, authority score) of a page, the more likely that the pages it points to (respectively, that point to it) have high authority scores (respectively, hub scores).

(c) (10 Points) According to the *Google Press Reviewer's Guide*, "Unlike other search engines, Google's advertising is precisely targeted and graphically designed to enhance a user's overall search experience." The claims of superior precision in targeting and enhanced user experience may or may not be self-serving, but the difference in graphic design is clear: Google displays textual ads against a white background and does not use banners, pop-ups, and other "aggressive" graphics to display ads. The two main components of the advertising program are Premium Sponsorship and AdWords. A Premium Sponsor buys a keyword, keyphrase, or category; when a user searches for this keyword, keyphrase, or category, the sponsor's URL appears first on the list of returned URLs, clearly identified as a "sponsored link." The AdWords program also allows an advertiser to associate itself with a keyword, keyphrase, or category; however, instead of causing the advertiser's URL to appear first on the list of returned URLs, AdWords allows the advertiser to monitor ad statistics and to display ad text next to users' search results.

Question 4.

(a) (4 Points) iv. all of the above

(b) (4 Points) False (See page 32 of 11/08/2001 lecture notes.)

(c) (10 Points) Online distributors of copyrighted works would like to maintain control over their digital content after it has been distributed to "untrusted" users. Web-service users would like to maintain control over their personal data after it has been transferred to "untrusted" servers. (See page 34 of 11/08/2001 lecture notes.)

(d) (7 Points) Copyright law