Homework 4

This assignment is due in class on December 4, 2001. Late homework will not be accepted.

Please write your name, email address, and date on the paper you turn in.

Question 1.
(a) (7 Points) What does Benkler mean by "peer production"?
(b) (6 Points) Give two examples of "open-source" software products that are widely used.
(c) (4 Points) For crawling eBay's website, the auction-data aggregator Bider's Edge was found guilty of
      i.   copyright infringemnet
      ii.   patent infringement
      iii.   trespass
      iv.   all of the above
(d) (4 Points) Which Supreme Court decision has been interpreted to mean that raw data are in the public domain and cannot be protected by copyright?
(e) (4 Points) Two necessary preconditions for the realization of Benkler's vision of "peer production" are
      i.   low-cost communication and low-cost access to information
      ii.   strong property rights and free markets
      iii.   strong property rights and corporate hierarchies
      iv.   information security and digital identity

Question 2.
(a) (5 Points) Franklin's article "How Internet Search Engines Work" (from www.HowStuffWorks.com) gives one reason that an owner of a web page may not want a spider to access the page and rank it for a search engine. If the page were part of a web-based game, the game might mistake the spider for a (human) player. Give another reason that an owner may not want to give a search engine's spider access to his or her web page.
(b) (5 Points) Web-page *meta-tags* can make search engines' jobs easier, *e.g.*, by making clear which of several meanings of a word on the page is the relevant one. Unfortunately, there is also a major downside to meta-tags. What is it?
(c) (3 Points) Before the emergence of the World-Wide-Web, there was a (much smaller) Internet-user community, and it created Internet-accessible information repositories, most often in the form of FTP sites. Give an example of a pre-WWW search engine that was used to find information on FTP sites.
(d) (8 Points) Google's effectiveness as a search engine stems partly from the number of pages it indexes (more than 1.6 billion) and partly from the way it uses the link structure of the web. In essence how does Google make use of links?
(e) (5 Points) What is the difference between a Google Web Search™ and a Google Site Search™?

Question 3.
(a) (5 Points) In Kleinberg's web-searching framework, what is "the abundance problem?"
(b) (10 Points) Also in Kleinberg's web-searching framework, what is a "hub," what is an "authority," and why are hub-authority relationships "mutually reinforcing?"
(c) (10 Points) How has advertising been incorporated into Google's business model? What distinguishes the way Google uses ads from the way many other websites use them?

Question 4.
(a) (4 Points) Which of the following security and privacy problems might arise in web services?
    - i.   eavesdropping on message traffic
    - ii.   impersonation (*i.e.*, lying about one's identity)
    - iii.   illegitimate use of personal data
    - iv.   all of the above
(b) (4 Points) True or False: The only entities that need to be reliably identified by web services are human users.
(c) (10 Points) Explain why the challenge faced by web-service users who are concerned about control and distribution of their personal information is analogous to the challenges faced by online distributors of digital content.
(d) (7 Points) To meet the challenge referred to in part (c), both web-service users and content distributors can benefit from security technologies such as encryption, digital signatures, P3P, and rights-management languages. However, content distributors have a powerful non-technical tool in their arsenals for which there is no analog in web-service users' arsenals. What is this non-technical tool?