Diddy	=	P. Diddy	=	Puff Daddy	=	Puffy	=	Sean Combs

Eminem	=	Slim Shady	=	Marshall Mathers

Preferred Term

Alternate Terms

Figure 4.2 Hip-hop authority files for Diddy and Eminem.

What's In a Word?

We all remember synonyms and homonyms from grade school. Let's take a moment to introduce some of the more esoteric terms for describing lexical relationships.

- *Homographs* are words with the same spelling but distinct meanings. For example, a ship has a bow, and actors bow after a performance. *Heterophones* have the same spelling but different meanings and different pronunciations, like *moped* (as in sulked) and *moped* (as in scooter).

- Words that are spelled the same but take on different meanings when capitalized are *capitonyms*, like *polish* and *Polish* or *nice* and *Nice*.

- *Hypernyms/hyponyms* indicate a more specific or more general relationship between words. For example, *duck* is a hypernym of *bird*, and *vehicle* is a hyponym of *car*.

- *Meronyms* describe a part of a whole indicated by another word. *Beak* is a meronym of *bird*.

- *Polysemes* are words with more than one similar meaning, like chair (something you sit on) and chair (person who leads a meeting).

If you want to dig more deeply into word meanings, have a look at WordNet, a lexical database maintained by Princeton University (http://wordnet.princeton.edu/).

USER-GENERATED CONTROLLED VOCABULARIES

Once again we'll look to LibraryThing, the innovative social cataloging Web site, for an example that mixes tagging with controlled vocabularies.

LibraryThing's "combine tags" feature lets users create an authority file for the site's tags. Any paying user can combine or separate two tags. For example, "science fiction" and "scifi" both refer to the same genre of futuristic literature; combining these tags makes them equivalent (see **Figure 4.3**).
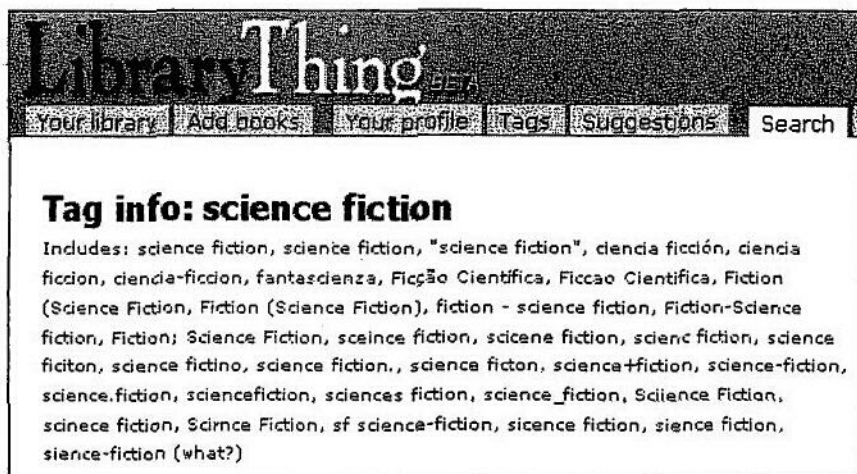

Figure 4.3 Combined tags for "science fiction" on LibraryThing.

LibraryThing also uses popularity to confer authority. When tags are combined, the most popular tag becomes the preferred term.

Combining tags is guided by a simple rule—combining should be used only to eliminate meaningless differences between two tags. This leads to situations where some tags that seem identical—"humor" and "humour"—are not combined because they are seen to hold important differences. And, indeed, the list of books tagged with "humor" includes more American authors such as David Sedaris, Scott "Dilbert" Adams, and Jon Stewart, although "humour" leans heavily toward Douglas Adams and Terry Pratchett.

In other cases, the subtle cultural differences between two seemingly synonymous tags have been ignored. For example, the tag "science fiction" and its Spanish equivalent "ciencia ficcion" have been combined.

The validity of these distinctions is left up to the LibraryThing community. By giving its users control over combining and separating tags, LibraryThing enables an ongoing conversation about differences between tags. If LibraryThing's users agree that "humor" and "humour" are essentially the same, they can be combined, and the system reflects the decision immediately.