We are more interested in median income or Star Trek than in national parks or in the vice president. We have more to say about the Summer Olympics than the Supreme Court, and we refer to hip hop more often than to the Soviet Union. Who is this "we"?

Well we are the contributors to the English Wikipedia, the on-line free content encyclopedia, with its 1.7 million pages, growing at a rate of more than 50,000 pages a month written by, at last check, about 3.4 million registered users, and, yes, a few automated computers programs generating pages for most of the US cities.

Wikipedia is not only an on-line resource you can browse with a computer, but it is also freely available for download from the Wikimedia foundation, the parent organization of Wikipedia, allowing anyone with a bit of computer savvy, or the help of their teenage neighbor, to install the whole collection of pages on a home computer, or for that matter, as some hobbyists have started doing, on a USB memory key.

With all the wiki pages on one's computer, and a bit of programming skills (see above teenager), one can perform some simple but enlightening experiments. How simple? Try this: have a program read each Wikipedia page, and keep track of the ten most frequent pairs of consecutive words in each page (you'll have to skip all the non interesting words, such as I, you, me, of, the, etc). Pairs of words are more representative than single words, and better capture the concepts expressed in a Wiki page than single words do.
Collect all the top-ten pairs found and count in how many wiki pages they appear. Then rank the pairs keeping at the top those appearing in the most pages. An interesting programming project for your high-school computer geek! Maybe not highly scientific, but a sure conversation starter at your next social dinner.

When the process is over, a look at the most frequent pairs reveals who this "we" of the first paragraph refers to. "We" is a chauvinistic, socio-economically minded American population who makes United States the most frequent pair of words in Wikipedia, appearing more than twice as often as the second runner up, New York, and more than three times as often as the third runner up, median income. Surely, one would argue that people in Britain and other English-speaking countries also contribute to Wikipedia. Yes, indeed. Who else would contribute to make Prime minister appear more often than democratic party, or republican party, both appearing more often than vice president?

By the way, democratic party appears almost twice as many times as its counterpart, with 6,160 pages versus 3,418, while vice president lurks in only 3,372 pages. But one shouldn't be too quick at faulting us for our chauvinism: the French and German also make United States the number one word pair in their respective language wikipedias!

The richness of information found in Wikipedia has not been unnoticed. A growing number of researchers are mining information from Wikipedia asking all kinds of surprising and interesting questions. The Wikimedia's research page lists a few of the questions currently investigated in universities: Is it still necessary to hire human experts (think bearded college professors) to argue on how to best categorize unstructured information describing our world, or are we smart enough to build in Wikipedia what is needed for the process? Can the information processing technology supporting
Wikipedia improve cooperation in education? Can Wikipedia help us learn more about the sociological aspects of cooperation and collaboration? How can the political parties better glean information from Wikipedia to create a more attractive platform to run on? What is the influence of bias in the information presented in Wikipedia? Is this type of democratic open content encyclopedia an example of participatory journalism? What is its influence on collaborative opinion forming and deliberation?

A system where information is freely posted, updated, and browsed by an entire community hides much to be explored. Even simple and simplistic experiments such as the one described here reveal a rich canvas of symbols and connections, hinting of what a community judges important, what it decides to record about itself, to emphasize, or to forget. History and popular culture are woven in the pages of Wikipedia. *Civil war* appears more often than *Vietnam war*. The term *African American* appears in nearly four times as many pages as *Native American*. The contributors of the encyclopedia, whom Wikipedia surveyed once as mainly young urban Caucasian males, write more often of *TV series*, *hip hop*, and *video games*, than topics relating to *North America* or *Roman Catholicism*.

So, what about the bottom of the list? Concepts that make the top-ten list of individual pages, but appear in just a handful of them? Lost in a sea of uninteresting words is this one: *sexual intimacy*. Dr. Ruth Weistheimer may have much to say about this.

*Note: More information on the methodology used to gather the statistics for this article can be found at http://cs.smith.edu/~thiebaut/wpl/. Note that this page will not be made public until the publication of the article.*