December 31, 2011
Defining Words, Without the Arbiters
By ANNE EISENBERG

TRADITIONAL print dictionaries have long enlisted lexicographers to scrutinize new words as they pop up, weighing their merits and eventually accepting some of them.

Not Wordnik, the vast online dictionary.

No modern-day Samuel Johnson or Noah Webster ponders each prospective entry there. Instead, automatic programs search the Internet, combing the texts of news feeds, archived broadcasts, the blogosphere, Twitter posts and dozens of other sources for the raw material of Wordnik citations, says Erin McKean, a founder of the company.

Then, when you search for a word, Wordnik shows the information it has found, with no editorial tinkering. Instead, readers get the full linguistic Monty.

“We don’t pre-select and pre-prune,” she said. “We show you what’s out there now. Then we let people decide whether to use a word or not.”

At one time, she was the head of the pruners, as principal editor of the New Oxford American Dictionary. She is also an author and columnist. (She wrote “On Language” columns for The New York Times as a substitute for William Safire.)

But Ms. McKean has chosen a different path at Wordnik. “Language changes every day, and the lexicographer should get out of the way,” she said. “You can type in anything, and we’ll show you what data we have.”

When readers ask about a word, Wordnik provides definitions on the left-hand side of the screen. But it is the example sentences, featured on the right-hand side, that are crucial to a reader’s understanding of a new term, she said.

“Dictionary definitions tend to be out of date or incomplete,” she said. “Our goal is to find examples on the Web that use the word so clearly that you can understand its meaning from reading the sentence.”

To do this, the site processes a vast reservoir of language, keeping tabs on more than six million words automatically, said Tony Tam, Wordnik’s vice president for engineering. “But the numbers change every second,” he said. “It’s not a static list.”

Where does all this text come from? “You’d be amazed how fast people write articles on the Web,” he said.

Wordnik does indeed fill a gap in the world of dictionaries, said William Kretzschmar, a professor at the University of Georgia and the former president of

the American Dialect Society. He provides American pronunciations for the new online Oxford English Dictionary.

“It takes time for words to get into the more formal, published dictionaries,” he said. “Wordnik is sensitive to what people are interested in now.”

Wordnik, which has raised $12.8 million in venture financing, plans to use its vast database of words and word associations at the site and in many business partnerships to be announced this year, said Joe Hyrkin, the president and C.E.O.

The products will be similar to recommendation engines, but more powerful, he said. If you like a particular book, for example, Wordnik can recommend a similar one based on its understanding of words used to describe the book, he said.

“We’re not just using tags and descriptors,” he said. “Our system understands and identifies matches at a concept level.”

The company is already providing many other word-based services, including one used on the Web site of The Times to define words in articles. Wordnik is also providing a financial glossary for SmartMoney.com.

Geoffrey Nunberg, a linguist at the School of Information at the University of California, Berkeley, who talks about language on “Fresh Air,” the NPR program, appreciates Wordnik’s breadth. “There’s a lot of useful information here,” he said. (He has also written commentaries on language for The Times.)

But he thinks that hands-on lexicographers could fine-tune the entries.

“The idea that you can pull lexicographers out of the loop and have an algorithm to mediate between me and the English language is goofy,” he said. “Without hand citations done by trained people, you get a mess.”

To illustrate his point, he noted flaws in a number of Wordnik’s definitions. The first definition of “davenport,” for instance, in three of the fives sources used by Wordnik is a kind of small writing desk. “It hasn’t meant that since Grandma was a girl,” he said.

People use a dictionary to find out what is correct, and what is incorrect, he said. “If I were a journalist looking to see if a word was being used correctly,” he said, “I wouldn’t put my eggs in the Wordnik basket.”

Mr. Tam of Wordnik said the site was constantly improving.

“We discover these words with algorithms, but they are never perfect,” he said. “We constantly have to make them better.”

WORDNIK and other new linguistic databases have come about largely because of the vast body of text on the Internet and improved algorithms for searching it, said Mark Liberman, a professor of linguistics at the University of Pennsylvania.

“We now have an archived shadow universe that contains almost everything we’ve written — trillions of pages of text of published books, and now, broadcast archives as well,” he said.

Readers could always tap this reservoir by looking up examples of new words in Google Books or Google News. “But what Wordnik is giving you is not as raw as a Google search of examples,” he said, “because Wordnik sorts and clusters the examples into different senses of the word.”

Another innovative database is at Brigham Young University, where Mark Davies, a professor of linguistics, has amassed a collection, the Corpus of Contemporary American English, 1990-2011, containing millions of words of running text from articles, transcripts of conversations, and other sources. The collection, which indexes 425 million words of text — 1,000 may be from a newspaper article, for example — has been built over the last three years. It shows how often a word is used, and the types of discourse in which it is found, be it conversational speech or academic prose.

The collection also lets users see words found near a new word. “If you want to see how a word is used and what it means, the best way is to look at words nearby,” Dr. Davies said. The words are called collocates. To look up collocates of “fantasy,” for example, see http://bit.ly/rImCuH.

Dictionary builders have come a long way since the days of Johnson and Webster, said Dr. Kretzschmar at the University of Georgia. “But we have computers,” he said. “We can manage this vast network of words online and appreciate it in ways that Johnson and Webster never could.”

E-mail: novelties@nytimes.com.

This article has been revised to reflect the following correction:

Correction: December 31, 2011

An earlier version of this article misspelled the given name of Wordnik’s chief executive. He is Joe Hyrkin, not Joel.