

BEYOND THE BUZZ ON WEB MINING

By Dominic Field, Carol Foley, Emmanuel Huet, Neal Rich, and Henri Salha

NEW TECHNOLOGIES THAT MINE social media for users' comments about brands, products, and services offer seemingly limitless potential for consumer insight. In contrast to the artificiality, limited sample sizes, and high costs of traditional market research, online comments are unfiltered, plentiful, and relatively inexpensive to collect. Scores of vendors promote their Web-mining products, claiming, according to one vendor, to provide "an unobstructed view into what consumers think." Swayed by the buzz, companies are scrambling to embrace this new approach to insight generation. However, a fact-based look at Web mining reveals that companies should be aware of some important caveats.

Although Web mining is a valuable and fast-improving tool, its specific nature limits its ability to generate truly useful consumer insight. It is a complement to rather than a replacement for traditional research methodologies. To avoid wasting your company's limited consumer-insight resources—and more important, to avoid

the dangerous misuse of Web-mining results—you must understand where Web mining works, where it doesn't, and what you can do to use it correctly.

Companies need guidelines to help them exploit Web mining's potential and avoid basic research flaws. With this in mind, The Boston Consulting Group's Center for Consumer Insight in partnership with the Leo Burnett advertising agency recently embarked on a research study of Web mining. The goal was to better gauge where and how Web mining helps generate insights that are truly valuable for decision making. Our study compared results achieved through Web mining with those from traditional in-person qualitative interviews and quantitative market research on a single topic—in this case, "healthy eating." We also carefully examined the nature and quality of the output from specific Web-mining technologies.

Our Web-mining study revealed four noteworthy facts about Web-mining results: they do not reflect a representative

sample, they are not exactly on point, they are not sufficiently deep, and they are not always precise and accurate.

Not a Representative Sample

More than 90 percent of the population in developed countries has Internet access, but online comments do not reflect this entire population. Our research, combined with Leo Burnett's long history of Web-mining experience, shows that just 14 percent of the population is responsible for 85 percent of consumer-generated online content about brands, products, and companies, as well as 52 percent of online-content consumption. This is a small, concentrated group of people who communicate primarily among themselves. (See Exhibit 1.) The posters are not only few in number, they also have a distinct profile and characteristic attitudes.

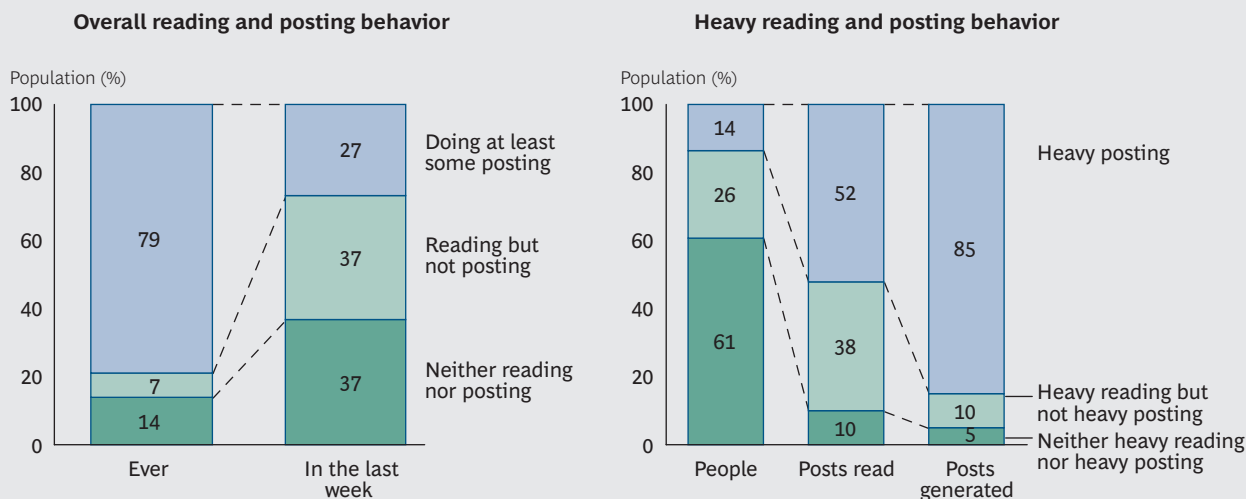
In our research on healthy eating, for example, the 6 percent of heavy posters who produce 80 percent of posts are significantly younger than the others and highly focused on appearing fit and trim. This is important for 70 percent of that heavy-posting minority but for only 30 percent of the rest of the population. You

may wish to capture the views of these highly engaged consumers in order to uncover unique perspectives or emerging signals, but you need to keep in mind that you cannot automatically generalize the insights gained through their posts to a broader group of potential customers. In our research, Web-mining output presented a much more positive overall picture of the healthy-eating topic, while our traditional qualitative-quantitative research program surfaced concerns, doubts, and frustrations about the topic.

Not Exactly on Point

Web mining is, in some ways, a double-edged sword. The beauty of the technique is that it is unguided. You can learn what consumers really want to talk about: topics they believe have social value for themselves or others. But the open nature of Web mining has a downside: you cannot steer it to where you most need consumers' input. Topics that are most critical to understanding preferences in a category may not be the topics that consumers choose to discuss. In our focus groups and in our quantitative survey research on healthy eating, "occasions" of eating (including day, time, and companions)

EXHIBIT 1 | Sample Bias Is a Fundamental Reality of Web Mining



Posters and readers are potentially the same people

Sources: "Beyond the Buzz," research conducted by The Boston Consulting Group and Leo Burnett, 2011; Leo Burnett experience; BCG analysis.
Note: Some totals do not equal 100 because of rounding.

surfaced as a critical issue, but Web mining fell short both in identifying eating occasions and in offering a basis for researching them effectively. In fact, the online conversation on healthy eating was vast and amorphous, with posts focusing on a wide but disorganized array of healthy-eating topics and a large dose of amateur advice. Although Web mining allows you to identify general themes and views, it provides limited information on how those views translate into actual purchase or usage behaviors. In most cases, you cannot follow up and ask online commenters for more information, nor can you learn about their specific personal context, habits, and spending in order to segment a market or estimate the size of the opportunity they represent.

Not Sufficiently Deep

For every thoughtful and articulate online comment, there are thousands of spare, poorly spelled, random, and even unintelligible posts. And such limited and often superficial expressions—consumers’ carelessly typed reactions to one another’s posts—are what you have to mine. By contrast, consider the richness and depth of well-run qualitative research through which in-person probing, observation, and decoding of nonverbal signals and projective techniques offer a finer understanding of behavior, attitudes, and motivations. Web mining lacks a mechanism for helping people better express what they mean.

Multivariate statistical techniques applied to well-designed quantitative survey data can reveal fundamental but hidden and often emotionally charged issues underlying dimensionality on a topic—issues Web mining would miss because of its reliance on conscious expression. For example, our survey analysis revealed that people organize healthy foods around a highly useful two-dimensional paradigm: physical versus emotional benefits of a food or an activity and its enjoyment versus deprivation. Web mining revealed no such organizational structure in the conversation. Due to Web mining’s unorganized nature and natural-content focus, we could identify

only a simplistic dichotomy based on weight loss versus healthy lifestyle. This was neither a new insight nor did it result in meaningful marketing implications.

Not Always Precise and Accurate

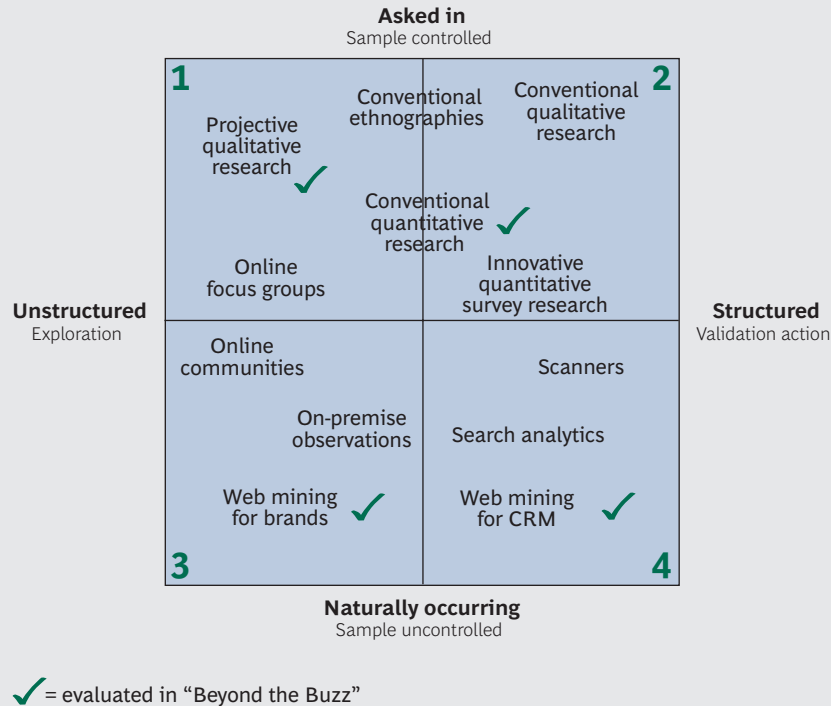
Although it has been improving, the quality of Web-mining tools and approaches still varies across many dimensions—from the essential removal of spam (which may comprise as much as 50 to 75 percent of the original data pull and as much as 25 to 50 percent of the “cleaned” data, depending on product category) to the accuracy of sentiment analysis. It is worth noting that the highest-performance techniques still include a significant human and judgment step in their protocols—driving up cost and timing but also improving quality. Furthermore, despite impressive progress in the analytics used to track the origins of online content, it is generally true that posters cannot be profiled. This severely limits the direct use of Web-mined insights for consumer segmentation or more extensive interpretation.

Despite these issues, we still believe that Web mining occupies a unique space in the research tool kit. (See Exhibit 2.) The “always on” and open nature of consumers’ expression on the Web is inimitable and powerful. The trick is to use Web mining selectively—recognizing its strengths as much as its limits—and to think carefully through the what, who, and how. As you incorporate these techniques at your company, you need to determine which categories are right for Web mining, its most powerful uses, who will own the insights, and how you will ensure high-quality results.

Which Categories Are Right for Web Mining?

The development of social-media platforms seems to make virtually any product, service category, or brand category a candidate for Web mining. But some categories clearly show greater levels of consumer engagement and corresponding

EXHIBIT 2 | Web Mining Occupies a Unique Space in the Research Tool Kit



Sources: “Beyond the Buzz,” research conducted by The Boston Consulting Group and Leo Burnett, 2011; Leo Burnett experience; BCG analysis.

Note: CRM = customer relationship management.

ly greater levels of social-media discussion. Within these categories, a rich base of available content somewhat reduces Web mining’s sample-bias problem, thanks to a more generalized posting pattern. For example, product categories that consumers find intrinsically rewarding to think and talk about—for example, travel, fashion, and video games—may lend themselves to Web mining better than, say, cleaning products or frozen food. Before engaging in Web-mining efforts, you should refine your understanding of your consumers’ posting and reading patterns—maybe as part of a general assessment of online behaviors in your category. It will help you determine the level of importance that should be assigned to Web-mining research and to enable your appropriate interpretation of its results.

For What Uses Will Web Mining Prove Most Powerful?

Considering its particular strengths, we see three main types of uses for Web mining.

- *Web mining adds a powerful exploratory lens.* Unguided and large-scale Web mining can help you hunt for topics and trends that may not emerge from a more controlled research study. In our test research on healthy eating, for example, Web mining brought to the surface a very specific topic: nutrition during pregnancy—a topic that had never come up in the spectrum of themes identified in our qualitative research. Furthermore, online posters may often be your most engaged, leading-edge consumers. What is on their minds? What early signals might you get from Web mining long before they bubble up into the mainstream?
- *Web mining is a critical tool for monitoring online conversations about your brand or category.* Posters are a minority of the population, but readers who don’t post likely represent the bulk of your consumer base and are being exposed to content from posters. Reviewing consumer-created content about your

brands, analyzing sources of content, and understanding specific “spaces of influence” online are all important. Web mining can, in this context, connect directly to your viral or word-of-mouth marketing efforts. In response to Web-mining input, some digitally advanced companies are also organizing themselves to provide the right information in the right places online or are inserting themselves directly into online conversations.

- *Web mining provides a window into real-time consumer responses to specific marketing actions and events.* Because Web mining is always on, you can gather feedback to a new PR campaign or capture the initial response to a new product. But keep in mind that posters are not representative. When Gap decided to drop its new logo in response to an Internet uproar, only 17 percent of the U.S. consumer population was even aware of the logo change (according to a survey commissioned by Ad Age). Was this a great example of “listening to the consumer” or an overreaction to a small and biased collection of vocal bloggers? It never hurts to confirm Web-mined insights through a wide range of representative data sources.

Who Owns the Insights?

Although traditional market research falls squarely in the functional home of the consumer insight team, social-media insights often lie in a no-man’s land. Companies outsource their brands’ Facebook conversations to their PR agencies, assign customer service reps to engage on Twitter with grumpy customers, and have internal digital-marketing teams monitoring social-media interactions. Such activities translate into “too many cooks” and, at times, result in partial and improper use of Web-mining insights.

In fact, Web-mining results are useful to multiple business functions, including marketing, corporate communications, and R&D. The goal should not be to limit the

use of Web mining or to centralize its execution, but rather to standardize vendors, policies, and methodologies. It is important to clarify the governance model up front, defining who owns and has access to the consumer conversation and who will take responsibility for gathering, synthesizing, and distributing Web-mined insights across the organization.

How Will You Ensure High-Quality Results?

Given the technical requirements of Web mining, it is generally both unavoidable and recommended to rely on external Web-mining vendors. But beware of sophisticated black boxes with many shiny buttons. It’s worth noting that there is an inverse correlation between having an attractive dashboard and producing real and accurate insight. The vendors that focus on the former generally produce less valuable insight because they devote insufficient effort to the time-consuming yet more reliable hand coding and human intervention that produce more accurate and insightful results. You must take a hard look under the hood before you select the tool you will use, mastering its possibilities and limits. The companies that leverage Web mining most successfully are those that ensure that they have at least the minimal level of in-house expertise to work with the vendors, challenge them, and evaluate the output themselves. These companies understand the mechanics—manual and automated—of the tools used, comprehend the limits of the automated tools, realize the importance of the human steps involved, and have established clear instruction manuals and processes, managing Web-mining studies with the same level of discipline they would bring to more traditional research.

Most important, we recommend that companies always validate Web-mining results through other data sources and research to ensure the accuracy of their results before using them as a basis for making costly and critical business decisions. When first adopting these techniques and evaluating the right strategy for

Web mining in your business, you may find that it makes sense to adopt a dual-methodology approach (Web mining mirrored by traditional research) on each key study for a sustained period of time.

WEB MINING IS a powerful tool that is evolving and improving and can serve as an excellent complement to traditional research techniques. Nonetheless, it remains best suited for specific kinds of jobs and should not be overused. Employ Web mining with reasonable care—as one component of a fully stocked consumer-insight tool kit.

About the Authors

Dominic Field is a partner and managing director in the Los Angeles office of The Boston Consulting Group. You may contact him by e-mail at field.dominic@bcg.com.

Carol Foley is an executive vice president and director of knowledge at Leo Burnett in Chicago. You may contact her at carol.foley@leoburnett.com.

Emmanuel Huet is a principal in BCG's Paris office. You may contact him by e-mail at huet.emmanuel@bcg.com.

Neal Rich is a principal in the firm's Chicago office. You may contact him by e-mail at rich.neal@bcg.com.

Henri Salha is a partner and managing director in BCG's Paris office. You may contact him by e-mail at salha.henri@bcg.com.

For Further Contact

If you would like to discuss this report, please contact one of the authors.

The Boston Consulting Group (BCG) is a global management consulting firm and the world's leading advisor on business strategy. We partner with clients in all sectors and regions to identify their highest-value opportunities, address their most critical challenges, and transform their businesses. Our customized approach combines deep insight into the dynamics of companies and markets with close collaboration at all levels of the client organization. This ensures that our clients achieve sustainable competitive advantage, build more capable organizations, and secure lasting results.

Founded in 1963, BCG is a private company with 74 offices in 42 countries. For more information, please visit www.bcg.com.

Leo Burnett Worldwide is one of the world's largest agency networks and the parent company of Leo Burnett and its marketing-services arm, Arc Worldwide. Leo Burnett, a HumanKind communications company, has a simple and singular approach: put a brand's purpose at the center of communications to truly connect with people. Leo Burnett, one of the most awarded creative-communications companies in the world, creates "acts, not just ads," for some of the world's most valuable brands including The Coca-Cola Company, Kellogg's, McDonald's, Hallmark, P&G, Allstate, and Nintendo.

© The Boston Consulting Group, Inc. 2011. All rights reserved.
11/11