INSIGHTS · CUSTOMER AND MARKET INSIGHT

Search Behaviour as Insight Into Customer Intent

How the searches people make on search engines and websites reveal their needs and intentions, how search data can be used to measure interest, spot trends and forecast demand, and the limitations and pitfalls that analysts need to keep in mind.

ModQu research article · 2,296 words · 22 references (APA 7th edition)

When people want something, they often search for it. A search query is a brief, unprompted statement of what someone wants to know, find or do, typed at the moment the need arises. Taken together, millions of queries form a record of public needs, interests and intentions that no survey could match in scale or immediacy. Search behaviour, whether on public search engines or on an organisation's own website and app, is therefore a rich source of customer insight. It is also easy to misread. This article explains what search data can reveal, how it has been used to measure interest and forecast behaviour, and what limitations need to be kept in mind.

What a search reveals

Broder (2002) analysed web searches and classified them into three types according to the searcher's goal. Navigational searches aim to reach a particular website, such as typing a company's name to find its homepage. Informational searches aim to find information about a topic. Transactional searches aim to do something, such as buy a product, download a file or book a service. Rose and Levinson (2004) developed a more detailed classification and found that a large share of searches were informational or aimed at obtaining resources, rather than navigational. Jansen et al. (2008) developed methods for classifying searches automatically by intent, and found that most queries were informational, with smaller shares navigational and transactional.

What a search reveals

Four types of search intent

Intent	Example query	What it reveals
Navigational	*“acme bank login”*	The brand is already in mind
Informational	*“how to fix a leaking tap”*	A need or problem, often early
Commercial investigation	*“best family car”, “A vs B”*	Active comparison of options
Transactional	*“buy running shoes size 9”*	Readiness to act

Based on Broder (2002) and Jansen et al. (2008). Diagram by ModQu ModQu.com

Figure 1. Search queries express different intentions. The words people use show how far along they are and what they need. Diagram by ModQu, based on Broder (2002) and Jansen et al. (2008).

For insight teams, the classification is useful because it shows where people are in their decision journey. Informational searches often reveal problems and needs before people have chosen a solution. Comparative searches, such as those including "best" or "versus", show active evaluation. Transactional searches show readiness to buy. Navigational searches for a brand show that it is already in mind.

Search queries are also short. Silverstein et al. (1999) analysed a very large log of queries to a search engine and found that most were very short, typically two or three words, and that most users looked only at the first page of results. Short queries are often ambiguous, which is one reason why interpreting search data requires care.

Search as a measure of interest

The volume of searches for a term can serve as a measure of public interest. Tools such as Google Trends report the relative popularity of search terms over time and between places. Mellon (2014) examined whether search data could measure which political issues the public considered important, and found that it worked well for some issues but poorly for others, depending on whether people's concern led them to search.

Search data has also been used to measure attitudes that people may be reluctant to state in surveys. Stephens-Davidowitz (2014) used the frequency of racially charged searches in different areas of the United States as a proxy for racial animus, and found that it predicted the vote share of a black presidential candidate beyond what was predicted by other factors. The anonymity of searching can make search data a window into concerns that respondents would not admit to an interviewer.

For businesses, search volumes show interest in categories, brands, features and problems. Du and Kamakura (2012) developed methods for identifying trends in large sets of search series, and showed how they could be used to spot emerging patterns in consumer interest. Du et al. (2015) found that trends in searches for product features, such as fuel economy for cars, could improve models of market response, because they captured shifts in what consumers cared about.

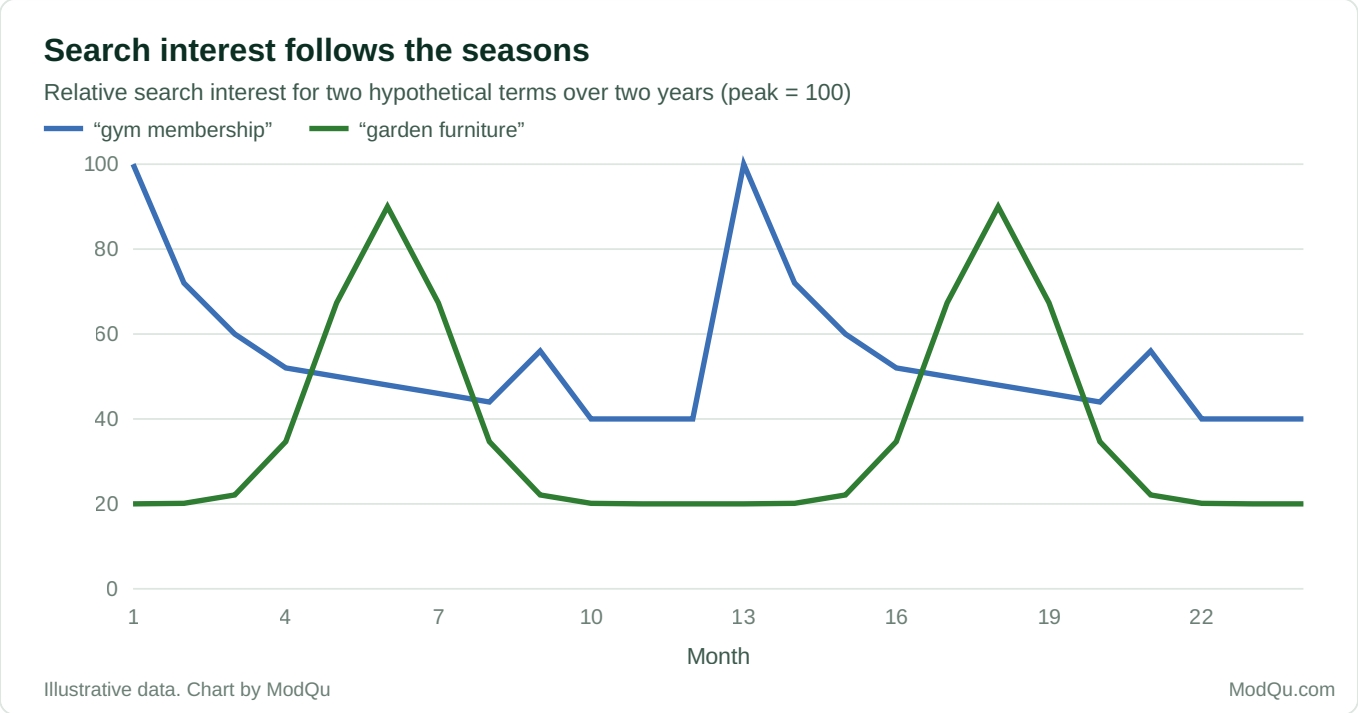


Figure 2. Search interest often follows seasonal patterns that show when needs arise. Illustrative data, chart by ModQu.

Search for forecasting

Because searches often precede action, search data can help predict behaviour. Varian (2014) described how methods from machine learning, combined with traditional econometrics, can help economists make use of large data sets such as search records, while warning that prediction and causal inference are different tasks. Choi and Varian (2012) showed that adding search data to simple forecasting models improved short-term predictions of economic indicators such as car sales, unemployment claims and travel, an approach they called "predicting the present". Vosen and Schmidt (2011) found that indicators based on search data outperformed traditional survey-based indicators in forecasting private consumption in the United States.

Goel et al. (2010) examined whether search volumes could predict the success of films, video games and songs, and found that they did predict outcomes, but often only slightly better than simple models based on other publicly available information. Their study is a useful reminder that search data adds most value when other information is scarce or slow to arrive.

The best-known cautionary tale concerns influenza. Ginsberg et al. (2009) found that searches for flu-related terms could estimate the level of influenza-like illness in different regions of the United States quickly and accurately, and the approach was turned into a public tool. Lazer et al. (2014) later showed that the tool had substantially overestimated flu levels in some seasons. They identified several reasons: the original model had picked search terms that happened to correlate with flu but were not related to it, changes to the search engine itself altered how people searched, and media coverage of flu prompted searches by people who were not ill. They called the problem "big data hubris", the assumption that large data sets can replace, rather than supplement, traditional data collection and analysis.

Jun et al. (2018) reviewed a decade of research using search trend data and described its growing use in many fields, from health and economics to marketing, alongside persistent concerns about how the data are sampled and normalised.

Search on your own website

Searches made on an organisation's own website or app are often even more useful than public search data, because they come from people already interacting with the organisation and can be linked to what they do next. Site search logs show what visitors are looking for, in their own words, and whether they find it. Searches that return no results, or that are followed by leaving the site, reveal unmet needs and gaps in content or products. Searches for terms that the organisation uses differently, such as customers searching for "cancel" when the organisation calls it "manage subscription", reveal mismatches in language.

Moe (2003) analysed navigation within an online store and found that visitors could be grouped into those buying directly, those searching deliberately for something to buy later, and those browsing, and that these groups behaved differently and responded differently to marketing. Bucklin and Sismeiro (2009) reviewed research on clickstream data, the record of how people move through websites, and described how it can be used to understand search, browsing and purchase behaviour. Joachims et al. (2005) studied how users click on search results and found that clicks are informative but biased: users tend to click on higher-ranked results whether or not they are more relevant, so click data reflect both preference and position.

Questions people ask

Some of the richest search insight comes not from volumes but from the exact wording of queries. Questions that begin with "how", "why", "can I" or "is it safe to" show what people are trying to understand and what worries them. Autocomplete suggestions and lists of related searches, which reflect what many other people have typed, give a quick view of the questions that surround a topic. A manufacturer of baby car seats, for example, might find that many searches combine its category with words such as "expiry", "second hand" or "fit in small car", each pointing to a concern that its products, packaging or website could address.

Analysing the language of queries also shows how customers name things. People often search using words different from those an industry uses, and the gap matters both for being found and for understanding how customers think about a category. Grouping queries by the problem they express, rather than by the product they mention, can reveal needs that cut across an organisation's product lines.

Search across the journey

Search data can be mapped onto the stages of a purchase journey. Early informational searches, about a problem or a general category, show the size and timing of emerging needs. Comparative searches show which brands are being considered together and which attributes people use to compare them. Branded and transactional searches show how many people have reached the point of choosing a particular supplier. Tracking the relative volume of searches at each stage over time, and for each brand, can show where a brand is gaining or losing ground: a brand that appears often in comparative searches but rarely in transactional ones may be considered but not chosen.

Search and advertising

Search behaviour is also the basis of a large advertising market, and research on search advertising shows how intent varies. Ghose and Yang (2009) found that searches containing retailer or brand names were associated with higher conversion rates than generic searches, consistent with the idea that branded searches indicate stronger purchase intent. Rutz and Bucklin (2011) found that generic searches could lead to later branded searches, so that early informational searches play a role in the journey even if they rarely lead directly to purchase.

But a search that precedes a purchase is not necessarily its cause. Blake et al. (2015) ran large field experiments and found that paying for search advertisements on a company's own brand name had little effect on sales, because people searching for the brand would have found it anyway. Search data show intent; they do not, on their own, show what changes it.

Limitations and pitfalls

Search data has several limitations that analysts should keep in mind.

Who searches. Not everyone searches, and those who do are not representative of the whole population. Some needs lead to searches; others are met in other ways.

Ambiguity. Short queries can mean many things. A search for "jaguar" may concern a car, an animal or a sports team. Interpreting search volumes requires checking what people actually meant, for example by examining related searches and the results people clicked.

Relative, not absolute, measures. Public trend tools usually report search interest relative to total searches and to the peak in the period, not absolute numbers, and they are based on samples. Small terms can show noisy patterns, and figures can differ slightly when downloaded at different times.

Changing systems. As Lazer et al. (2014) showed, search engines change their algorithms, autocomplete suggestions and interfaces, which change how people search. Patterns that held in the past may not continue.

Media effects. News coverage can prompt searches by people who are curious rather than interested in buying. Spikes in search volume after a news story may not reflect a change in demand.

Overfitting. With thousands of possible search terms, some will correlate with any outcome by chance. Choosing terms on the basis of theory and checking predictions on new data protects against this.

Privacy

Search data are personal. Individual queries can reveal health conditions, financial difficulties, relationships and other sensitive matters, and a sequence of queries can identify the person who made them. Public trend tools address this by reporting only aggregated, relative figures. Organisations analysing their own site search logs should do the same: aggregate queries before analysis, remove personal details that people type into search boxes, such as names, account numbers or email addresses, and use the data only for purposes compatible with what users would reasonably expect. The ICC/ESOMAR (2016) code for research and data analytics requires that personal data be protected and that people not be harmed by the use of their data, principles that apply as much to search logs as to surveys.

Using search insight well

Several practices help. Start with a clear question about a need, interest or behaviour, and choose search terms that plausibly reflect it. Classify queries by intent, so that informational, comparative and transactional searches are not mixed together. Read samples of actual queries to understand what people mean. Compare search trends with other sources, such as sales, surveys and social media, before drawing conclusions. Use site search data to find unmet needs and language mismatches. And treat search-based forecasts as one input among several, validated on data that were not used to build them.

A short example shows how these practices combine. A hypothetical home insurer notices that searches combining its category with "flood" have risen sharply in several regions over two years, well beyond the short spikes that follow news of storms. Reading the queries shows that many ask whether standard policies cover flooding and how to check a property's risk. Its own site search logs show the same questions, often followed by visitors leaving without a quote. The insurer adds clear flood-cover information and a postcode risk checker to its site, and tracks whether quotes from those regions rise. The search data did not tell the insurer what to do, but it showed a growing worry, where it was concentrated and the exact questions customers needed answered.

Key takeaways

Search queries are brief, unprompted statements of what people want, made at the moment of need. They reveal intent, from informational searches early in a journey to transactional searches close to purchase, and their volumes can measure interest, reveal seasonal patterns and help forecast behaviour. Site search data show what customers look for and fail to find. But search data come from a particular population, are often ambiguous, are reported in relative terms and can be distorted by changes in search engines and by news coverage. Used with care, and combined with other evidence, search behaviour is one of the most direct windows into customer intent.

References

Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica, 83*(1), 155–174. https://doi.org/10.3982/ECTA12423

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum, 36*(2), 3–10. https://doi.org/10.1145/792550.792552

Bucklin, R. E., & Sismeiro, C. (2009). Click here for Internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing, 23*(1), 35–48. https://doi.org/10.1016/j.intmar.2008.10.004

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*(S1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Du, R. Y., & Kamakura, W. A. (2012). Quantitative trendspotting. *Journal of Marketing Research, 49*(4), 514–536. https://doi.org/10.1509/jmr.10.0167

Du, R. Y., Hu, Y., & Damangir, S. (2015). Leveraging trends in online searches for product features in market response modeling. *Journal of Marketing, 79*(1), 29–43. https://doi.org/10.1509/jm.12.0459

Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science, 55*(10), 1605–1622. https://doi.org/10.1287/mnsc.1090.1054

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012–1014. https://doi.org/10.1038/nature07634

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences, 107*(41), 17486–17490. https://doi.org/10.1073/pnas.1005962107

ICC/ESOMAR. (2016). *ICC/ESOMAR international code on market, opinion and social research and data analytics*. ESOMAR.

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management, 44*(3), 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 154–161). Association for Computing Machinery. https://doi.org/10.1145/1076034.1076063

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change, 130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Mellon, J. (2014). Internet search data and issue salience: The properties of Google Trends as a measure of issue salience. *Journal of Elections, Public Opinion and Parties*, *24*(1), 45–72. https://doi.org/10.1080/17457289.2013.846346

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology*, *13*(1–2), 29–39. https://doi.org/10.1207/S15327663JCP13-1&2_03

Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 13–19). Association for Computing Machinery. https://doi.org/10.1145/988672.988675

Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, *48*(1), 87–102. https://doi.org/10.1509/jmkr.48.1.87

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, *33*(1), 6–12. https://doi.org/10.1145/331403.331405

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, *118*, 26–40. https://doi.org/10.1016/j.jpubeco.2014.04.010

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578. https://doi.org/10.1002/for.1213