What Do You Need to Know to Stay Competitive in Mining the Social Web?

- **Skillful mining of Big Data** has become a critical component of business success.
- Analysis of social media data crosses **geography, language, and culture**.
- With users of social networks growing exponentially, the worldwide quantity of **digital data doubles every year**.
- Going **beyond sentiment analysis**, companies increasingly seek predictive insights into the influences, thoughts, emotions, and ideas that drive consumer buying decisions.
- Mining the social web is expanding to include **images, audio, video, and emoji**.

Does your solution comprehensively monitor social data to track your customers’ brands? Can your customers use social media to discern what people like and dislike about their products? Are they able to evaluate marketing campaigns in different countries?

The essential key to all of it—analyzing customer experiences, market research, consumer insights, digital analytics, and media measurement—is high-quality, multilingual **Text Analytics**. Unlike years past when machine translation was the only option, today’s cutting edge solution applies **linguistic analysis** to structured and unstructured text (news, reviews, blogs, tweets, and posts), so each word is understood in its **native** context. It’s the only way to deliver results that are not skewed by subtle errors in slang, syntax, or spelling.

The tools your customers need to successfully mine any data feed exist. And in an increasingly crowded marketplace your software must:
- Provide high quality results across ALL the languages your customers want
- Offer robust features, functionality, and scaling.
- Integrate easily into existing infrastructure

Your solution should also meet both your customers’ current needs and position them well for future growth.

Nine questions to ask about your software on the next page
Build vs. Buy

Your best solution depends upon what your customers want and how fast they need it.

Thoroughly research all options because pros and cons vary depending on your situation.

Consider the engineering costs associated with integration.
- Will you need resources and support to troubleshoot?
- Is there sufficient natural language processing expertise in-house?
- How critical is text analytics to your business value?

Do you need more than one Language?
If you are shopping for multilingual capability, it adds significant complexity, requiring expert knowledge in each language.

Ultimately, your choice may rest on the available in-house resources and your specific time-to-market requirements.

1. How does your text analytics handle short text, like Tweets?
Tweets have traditionally been more difficult to analyze because there is less context to work from, and they often include slang, abbreviations, and emoticons. But many solutions are now capable of excellent Twitter analysis—identifying the language and finding mentions of people, places, and companies.

2. Can my system analyze text in industries where jargon and specialized vocabulary exist?
If your customers use domain-specific vocabulary, they need analytics that can be trained for greater accuracy over time. So, look for text analytics that work right out of the box, but also can adapt to meet the specific and evolving requirements of any domain.

3. How do I guarantee high quality results across all languages?
Any analytics you build is only as good as the linguistic analysis foundation it’s based on. Machine translation is old school. You need a solution where each language is understood natively. This linguistic approach does not find related words based on how they appear but rather, what each word means within its written context. It’s the best way currently available to ensure all data is interpreted correctly.

This is also a reason you may prefer a company whose core competency is text analytics rather than one where analyzing Big Data is just one aspect of a wider suite of products.

4. How many languages should my social web mining solution accommodate?
To best position your company for future growth while minimizing both integration headaches and time-to-market, the short answer is as many as you are likely to need. But rather than adding languages piecemeal over time, you may be better served by a vendor known for the quality of its multilingual capability, across many languages.

5. How well does your product respond to the idiosyncrasies of search in different languages?
Comprehensive and reliable search results depend on native understanding of each language. Minor variations in spelling (color vs. colour) and characters (Tschüss vs. Tschueess) exist in all languages. The complexities of Chinese, Arabic, and Japanese pose greater challenges. Can your search engine accommodate all these variations?

6. How well does your product track names in multiple languages?
Because many brands are now recognized internationally, multilingual capability has become critical. Mentions of your brand need to be correctly identified across languages and scripts, regardless of slang, usage, spacing, or misspellings. For example: “KrispyKreme”, “Crispy Cream”, “crispy crème” and “クリスピー・クリーム” all refer to the same company.

7. Can your product distinguish between Samsung and Samsung Electronics?
To answer YES, you need a solution that uses entity extraction based on context, rather than exact match. And, once an entity is identified, you then need reliable entity resolution to accurately match the identified entity to a real-world thing, distinguishing it from similarly named entities.

8. Will your solution help me identify new competitors?
Advanced text analytics CAN discover relationships between one brand and others, which is how your customers will discover new and upcoming brands as they come to market. Any time a new entity is referenced, they should be able to flag and monitor it for the future.

9. Which solution is best at sentiment analysis?
A relatively new aspect of text analytics and the current darling of social web mining, sentiment analysis is improving all the time as software gets better at understanding the subtleties of context, sarcasm, and human error. The best solution will be one with algorithms that can be improved over time to suit your needs.