

Basis looks good to exploit changing text-analysis market in any language

Analyst: [Nick Patience](#)

Sector: [Enterprise Software](#)

Date: 12 Jun 2008

Basis Technology is one of the oldest text analysis companies around. Because much of its business comes from government deals – many in the intelligence sector – or OEM deals, many of which are confidential, the nature of its business tends to be low profile. But with its main rival now part of the **SAP** behemoth, its status as a pure-play provider of multilingual text analysis software looks like a good place to be. However, the headline of our [last report](#) indicated it was looking for expansion beyond its core OEM and government markets, but that hasn't really happened yet.

The 451 Take

Basis is in a good spot right now – profitable, not needing any external investment (which is the best time to look for it should it want to) and standing almost alone as a provider of multilingual text analysis software. Why it hasn't managed to get more OEM deals with nonsearch companies is a slight worry though, as it will increasingly come up against much bigger and better-funded rivals, although that is awhile away.

Context

Cambridge, Massachusetts-based Basis Technology specializes in multilingual information retrieval and entity extraction. Founded in 1995, the company has been profitable since 2003. Early investors included **Accenture Technology Ventures, Amazon.com, In-Q-Tel** and **Kyocera Goldman Sachs Venture Capital**. But the last funding round – an undisclosed amount from In-Q-Tel – was back in 2004, and Basis is not currently seeking funding. However, the company maintains contacts with potential investors and is interested in exploring ways to use further investment. The company has hired 10 since May 2007 and now employs 65. It is hiring in most areas, including engineering, sales and product management.

In January, Basis won a long-running lawsuit with investor Amazon.com over management and consulting services provided during the creation of Amazon's Japanese subsidiary and its e-commerce website. Amazon.com has since left the Basis board of directors and is no longer a preferred stock shareholder.

Products

Version 6 of the flagship Rosette Linguistics Platform (RLP) was released in February. This is the product that contains all the various language identification, base linguistics and entity extractors that are licensed for use in other vendors' software but also underpin Basis' other products. New languages added in version 6 are Farsi and Urdu. RLP can now enable searching and text analysis natively in 18 languages and can identify 54 languages in all. The base linguistics include tokenization, stemming, part-of-speech tagging, noun phrase extraction and sentence boundary detection. Not every linguistic analysis technique is available for every language, but it is a pretty broad set nonetheless.

Rosette Name Indexer and Name Translator are more recently added parts of RLP (their numbering is up to 2.0 rather than 6.0 for the rest of it). Name Indexer, which used to be called Name Matcher, sports the **Lucene** set of open source search engine libraries under the hood. It can take a name, spelled in any of the languages it supports, and match it to a name in a database – even if the name is spelled incorrectly or in a different language or script. For instance, if someone, upon hearing the name of the president of China for the first time, thought his name was spelled 'Hoo Jintow,' it promises to be able to match that to the correct name in normalized English form (Hu Jintao) or the actual Chinese characters in simplified Chinese. Basis has sped up Name Indexer's algorithms, which requires fuzzy lookups that are usually very time consuming.

The uses for such software in counter-terrorism are fairly widespread. An intelligence agent on the ground might hear chatter about a certain person and need to take an educated guess about the person's name. The US government's databases such as the Office of

Foreign Asset Control of the **US Department of the Treasury** or Political Exposed Person are all in English and often contain dirty data as names have been transliterated from their original language.

Name Translator is a tool to find the most accurate English translation of a non-English name, plus Basis supports all six of the US government's transliteration standards. Basis has been working on its entity extractor, adding nationality and region as new entities.

Sales

In May 2007, Basis reported more than 40 government and 200 nongovernment customers. Most of the latter are OEMs, including **FAST Search and Transfer, Endeca, Autonomy Corp, Google, MSN** and **Yahoo**. Since then it has won 10 more government clients as well as **Kroll Ontrack**, an added risk, e-discovery and computer forensics shop, on the nongovernment side.

Roughly half of Basis' government implementations involve using the products out-of-the-box, and the other half involve customization. The main government opportunities are document/media and cell phone exploitation (in other words, reading documents and listening to phone calls) as well as selling the desktop suites for use by intelligence analysts.

New partnerships with search and database companies are set to be announced this year as well. Basis reports a measured but cumulative response from former **Inxight** OEM customers in the wake of the **Business Objects** and SAP acquisitions. Overall sales are 30% direct, 70% indirect (including the OEMs). Deals range from \$50,000 for royalties up to multimillion-dollar deals on the high end, with an average size of \$250,000-300,000.

Current target markets include compliance and fraud detection in the general enterprise market as well as forensics for government and e-discovery. The company has found forensics for law enforcement a bit tougher, with longer sales cycles and small deal sizes, so it is focusing more on forensics in the government and e-discovery markets.

Competition

Basis' main competitor used to be Inxight Software, but since the acquisition of that company by Business Objects in May, and the latter's acquisition by SAP in October 2007, things have been thrown into a state of flux, to an extent. That's not to say SAP is abandoning Inxight's extensive OEM and federal government, because it is not – what was Inxight SmartDiscovery, is now called Business Objects Text Analysis. But inevitably, internal integration takes up some of the Inxight team's time that was previously used competing against Basis and others.

Teragram was a less significant competitor to Basis prior to its acquisition by **SAS Institute** in March. SAS bought that company partly because its Text Miner product was reliant on Inxight technology, and it wanted to own its own text analysis product, although it will be a while before it's integrated into SAS. Meanwhile, SAS is maintaining Teragram as a distinct brand to help it compete as an OEM provider. Longtime SAS rival **SPSS** has long had its own text analysis software but doesn't really compete in the OEM market in which Basis does so much of its business.

Nstein Technologies used to be a closer competitor but has gone off in a different direction and is totally focused on large publishing companies, having combined text analysis with Web content and digital asset management. **Temis** is another one Basis may still occasionally see in deals.

Attensity is a competitor, but is also often combined with Basis as part of a broader package being put together by systems integrators in government deals. But in competitive deals, if non-English languages are needed, it can't offer true multilingual parsing and analysis of text. Attensity translates the text into English and then analyzes it, rather than analyzing it in the original language.

SWOT analysis

Strengths

Basis has a long history and strong technology in the area of multilingual text analysis. It is a key provider to many other software firms.

Weaknesses

Basis has not managed to expand much beyond government and OEM markets; we'd have thought in the OEM field more nonsearch companies would have bitten by now. The headline of our last report indicated it was looking for such expansion, but it hasn't really happened.

Opportunities

Inxight hasn't left the market, but it is certainly spending some of its time focused on internal integrations within SAP, which must leave gaps in the market for Basis to exploit.

Threats

SAS and SAP represent the future competitive threats to Basis. It will be going up against competitors with far deeper products than the standalone Teragram and Inxight.