

Basis Tech looks beyond core government and OEM markets for new opportunities

Analyst: [Katey Wood](#), [Nick Patience](#)

Date: 22 Jul 2009

In the wake of the acquisitions of two of its rivals, **Inxight** and **Teragram**, **Basis Technology** has been steadily accumulating OEM deals for its multilingual text analysis. Its historic strengths in government deals for its intelligence and forensics software have remained firm, with business picking up under a new administration. The last piece of the puzzle is attracting more commercial customers, and the company is making plans for that as well.

The 451 Take

Basis is one of the oldest independent text-analysis companies still around – and the only pure play we're aware of with such extensive multilingual capabilities. The company has tackled government work with Middle Eastern and Far Eastern languages, and now seems to be making up ground with the more approachable European ones in its core linguistics platform. This will go a long way toward increasing its commercial appeal, as will its upcoming native support of Lucene, the open source search library many companies are now turning to as the foundation for their search platforms. The increasing internationalization of business also points to a bright future for the multilingual company, although with it will doubtless come more competitors.

Context

Basis has a long history with government intelligence, including a 2004 funding round from **In-Q-Tel** for an undisclosed amount. Government – especially the intelligence community – is, after all, a sector that knows exactly what to do with multilingual entity extraction as it looks for people, locations and relationships across multiple languages.

However, Basis is now focusing extra attention on the commercial space, where there is still a certain amount of educating to do about the potential for its technology, but it feels there is also demand not yet being met. The company began its push with the addition of a VP of commercial sales, Don Reigrod, in March. Reigrod previously worked as Senior Business Development Manager at **SAP**, general manager of the OEM division at Inxight, and VP of OEM and Channel Sales at **Fast Search and Transfer**. He also worked at **Inso**, the company behind the frequently OEMed Outside In technology, which it sold to **Stellent** in 2000, which was subsequently acquired by **Oracle** in 2006.

The historically engineer-heavy company also intends to double its marketing team and add more sales engineers. Since last year, employee numbers have held fast at 65.

Products

Releases from 2008 have expanded on Basis' existing offerings, including the addition of Russian to its entity-extraction languages. In Q4 2008, new name-matching capabilities for Latin languages were added as part of the Name Indexer module of its Rosette Linguistics Platform (RLP). Name Indexer takes a given non-English name and matches it against a database of thousands of names in different spellings and languages to identify it correctly, even if the name is misspelled or spelled phonetically. European languages are being added as part of the effort to further commercialize the product – the initial support for only Middle Eastern and Far Eastern languages was aimed more at government usage.

The roadmap for Q3 2009 includes linguistics modules for users of the open source **Lucene** search library and Solr, an open source enterprise search server. Basis had an existing connector for Lucene, which was refreshed last fall, but there has been more demand recently, with some customers requesting a cleaner integration. The Rosette Linguistics Platform is a C++ engine, while Lucene and Solr are both pure java tools. Also in Q3, Basis will be adding support for the Android and **Symbian** mobile operating systems as a result of interest from Japanese clients. So far, there are no plans for integration with the iPhone.

Future developments include the next release of RLP, which is expected before the end of the year. Improvements have been made to the entity extractor for better speed and accuracy, and it will include support for Pashto and Dari, both languages spoken in Afghanistan.

Sales

In 2009 came several new OEM deals with e-discovery companies for multilingual document processing, among them Ipro, **Clearwell Systems** and **Daticon/EED**. All of these companies sell tools used in response to litigation for sifting through and culling terabytes of potential evidence to find responsive documents for attorney review. Basis also won a deal for Chinese and Japanese search of **Mimosa's** NearPoint archive, often used for compliance. Interest in these areas is expected to grow internationally as the US standards for compliance and discoverability of electronically stored information begin extending across Europe and the rest of the non-English-speaking world. Other recent OEM deals have been made with XML server vendor **Mark Logic** and search vendor **Attivio**.

Interest in Lucene-based search has been another growth area, as evidenced by the new modules coming out this year. Several plans are in the works for enterprise-wide deals, and the company has been in talks with open source search consultant **Lucid Imagination** regarding a future partnership.

Altogether, Basis has about 50 government and 200 nongovernment clients under active maintenance, with the majority of those coming from direct sales. With commercial business slow during the recession, the gap has been filled recently by government deals, which have picked up since the new administration took office. Several large licenses have been sold in 2009 for document exploitation and triage, or the reading of adversary documents and ascertaining of their usefulness. More new projects and potential new deals are expected after the end of the federal fiscal year in September.

Overall business is still centered largely in North America, with about 5-8% of total revenue coming from Japan and a cluster of deals in Saudi Arabia. Average deal size is around \$250,000-300,000.

Competition

Basis had a worthy OEM adversary in Inxight prior to its acquisition in May 2007 by **Business Objects**, which was subsequently bought by SAP that September. The brand was subsumed into what is now Business Objects Text Analysis – we'll be reporting more on that integration very soon. Teragram is another company with extensive multilingual text-analytics capabilities. It was picked up by **SAS Institute** in March 2008, in part due to the purchase of Inxight, from which SAS licensed text-analysis technology. SAS is running Teragram under its existing brand name and is gradually integrating the Teragram technology into its own tools, as well as continuing to sell on an OEM basis. However, in the wake of these acquisitions, Basis is looking increasingly more like the go-to OEM vendor for multilingual text analysis among vendors that may compete with SAP or SAS in some way.

Attensity Group is another text-analysis vendor, but it previously focused its considerable text-analysis assets on voice-of-the-customer applications – and only with English parsing capabilities – and not so much on the OEM opportunity. However, this could change with its April rollup into the Attensity Group holding company along with two German text-analysis and content management vendors. New European blood could mean more languages for Attensity, and it has already announced applications for risk and compliance, research and discovery in corporate and legal processes, and intelligence analysis for government agencies. Basis could be seeing more of it in the future.

Linguamatics offers natural language processing (NLP)-based text-analysis software but focuses on markets such as life sciences, pharmaceuticals and healthcare. **Content Analyst** is another company racking up e-discovery OEMs with its own text analysis; however, its technology is statistically based rather than linguistic, and it learns through training documents rather than natively identifying languages like Basis. The statistical and NLP approaches have different strengths and weaknesses.

SWOT analysis

Strengths	Weaknesses
Basis has a strong foundation of multilingual text analysis, including a solid network of partnerships and good traction in the government sector.	The company is still setting up a sales and marketing framework to increase the commercial adoption of its products.
Opportunities	Threats
Its headway in e-discovery and compliance could foster more international deals as countries around the world adopt more stringent regulations for business and legal procedures.	Yet more consolidation in text analysis could create more competition for the company – once integrations have been completed.