



COLTEC: Gaining corporate competitive advantage through effective data mining

Business intelligence and information technology have recently joined forces to provide companies with a powerful new methodology with which to gain competitive advantage and make their customer relationship management (CRM) not only more effective, but more valuable.

Today, corporations the world over have rivers of information flowing into them every day from customers across the globe. Information that could help provide keen insight into consumer trends, purchasing preferences, product development, and much more. The challenge is to find that information within all that data. And while the previous sentence may seem nothing more than just a clever play on words, it is a simple but problematic truth. Corporate databases are overflowing with data that remains unanalyzed, its value unrealized.

As a result, corporations are investing heavily in analytical software and discovery applications that enable them to extract meaningful data from massive amounts of information. This information exists in two parallel universes – structured and unstructured data. Structured data is everything that is used to fill data fields: names, addresses, sizes, amounts, product numbers, and colors. Unstructured data is primarily text: e-mails, call center logs, customer complaints, sales rep reports, project notes, white papers, and the myriad other documents that don't fit the narrow confines of a specified data field.

The importance of unstructured data

Why is all this data that doesn't quite fit pre-defined categories so important? Because it provides the context – the insight – into the real story. For example, a warranty work form filled out by a technician might contain important

information such as what parts and how many were replaced for a particular product, but it doesn't provide the answer as to why those parts failed. That information could be contained in the technician's notes, along with insight into how the problem might be remedied. There's no way to discover that information unless some kind of software is used to "mine" those valuable nuggets from the heap of unrefined data collected over time by manufacturers and other companies.

Unstructured data may hold the key to the development of the next new product. Perhaps sales reps have been having a difficult time selling a particular product size or set of functions and note that in their sales reports. In the meantime, a development team is designing the next generation of that very product, investing hundred of thousands of R&D dollars into the project without realizing the size and functionality they favor is, in fact, not what the market really wants. The ability to "tease" that information out of the sales data would be crucial for saving the company considerable time and money in the development process.

Today, this unstructured data is being collected internationally. As we become a true global economy with new markets for new products springing up all over the world, increasingly those markets are being developed in Arabic-speaking regions, a language spoken by over 300 million people in over 20 countries.

Extracting value from Arabic information

Data analysis and information extraction are two of the most powerful tools for exploiting business intelligence, as corporations maneuver to gain share in new markets. COLTEC is a leading provider of Arabic language tools specifically designed to help companies process and manage documents, files, and notes to find relevant information, analyze context, and create valuable, indexed databases from massive quantities of structured and unstructured data.

With a growing shortage of trained, experienced Arabic translators and analysts available worldwide, our software fills a critical gap that enables companies and organizations to effectively manage, analyze, and act upon information in many of the world's most economically active and important regions.





Give your application added value and a bigger market with our products

If you're a software developer, you can add instant value and expand your potential customer base by incorporating our Arabic language products into your application. With one of our software development kits (SDK), you can give your content management system Arabic functionality not available with western language software. Our Proofing Tools are ideally suited for virtually any application with text processing capabilities – add Arabic spell and grammar checking and a thesaurus and dictionary for comprehensive, effective Arabic language processing. Licensing our products for use in your applications opens up one of the most dynamic and growing markets in the world for your company and your software – the Middle East.

What makes COLTEC different – and better.

We are the only Arabic language processing software developer in the world with a truly mature foundation in NLP, a foundation that has been under continuous development for over ten years. Staffed with a core of experienced, highly trained multi-language scientists, COLTEC's fundamental advantage is our patent-pending Natural Language Processing (NLP) model. Developed by our founder, Dr. Taghride Anbar, this revolutionary model is the basis for processing, analyzing, and managing Arabic as a linguistic system suitable for computer manipulation. This powerful, fundamental methodology is the very core of our business – it turns complex linguistic rules into flexible, logical communication tools designed specifically for computer use.

Our strengths include:

- Strong linguistic foundation and infrastructure
- Native Arabic-speaking developers
- Strong balance of linguistic and statistical expertise
- Ability to "identify, analyze, and even generate" Arabic words
- Processing speeds multiple times faster than our competitors, for example our Word Identification tool is able to process over 600,000 words/second
- A fully linguistically processed balanced corpus of Modern Standard Arabic (MSA) that includes over 150 million words
- A Lexical Database that uses proprietary compression techniques that result in a total size smaller than 3 MB

As a result, we are uniquely positioned to help companies worldwide process, analyze and manage Arabic data more efficiently and effectively than any other company or software.

We offer a range of products ideally suited to provide Arabic data management and analysis capabilities to corporate customers.

These include:

Arabic Proofing Tools offer the world's most comprehensive set of integrated Arabic proofing modules, designed to provide the user with everything necessary to confidently work with Arabic text, including a grammar checker, spelling checker, thesaurus, and dictionary. These tools are so intuitive and effective Microsoft has been using them exclusively in Office applications since 1997.

Ideally suited for a wide range of users – consumer, academic, and professional – our proofing tools include the most reliable, comprehensive Arabic thesaurus available; a multi-language dictionary for fast, bi-directional translation between Arabic, English, and French; a word corrector that actually spell-checks and corrects all types of word errors; and a grammar checker that automatically corrects grammatical and typographical errors.

Arabic Search Plug-In (ASPI) specifically designed to help Arabic & non-Arabic speakers search Arabic data. It takes into account the unique writing features of Arabic, as well as issues related to Ambiguity & affixations. No other Arabic search tool is based on a language model as subtle and sophisticated as ASPI.

ASPI is an ideal tool for discovering items of interest or information on a particular subject when users "don't know what they don't know". For example, a consumer looking to find a particular product in a popular color might not realize that "red" can also include maroon, scarlet, and burgundy.

Arabic Entity Extraction Tool (ANEE) uses advanced discovery search methodologies to extract critical information from large amounts of structured and unstructured data – unlike search tools, it analyzes concept and context and delivers results based on meaning. For example, a researcher or student wishing to discover any documents regarding international oceanographic research and the Mediterranean Sea can use ANEE to uncover key data linking government agencies, private research firms, and the world's oceans, even if names can be represented in a variety of ways, such as shortened versions, abbreviations, or even different spellings.

Word Conversion (WORDCON) accurately and reliably converts Arabic characters into Latin ones and vice versa, enabling crucial information, once found, to be converted for more thorough examination and analysis. Our proprietary linguistic foundation ensures correct conversion based on meaning and context in a language where over 60% of the words have more than one meaning or spelling. This can be especially useful for, say, non-Arabic speakers who need to make a presentation in Arabic and would find it helpful to have the words converted to western characters for proper sounding out.

