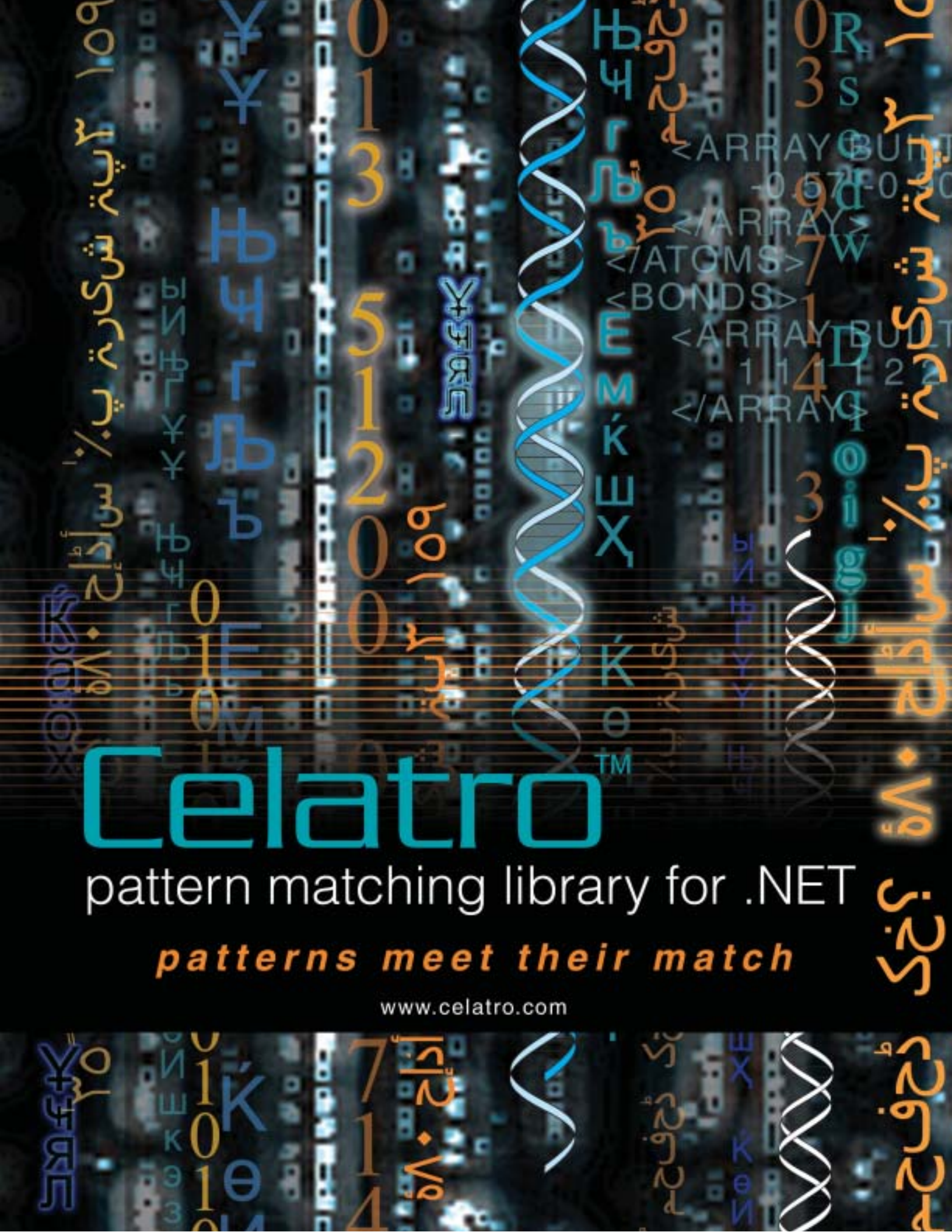

Celatro™
pattern matching library for .NET
patterns meet their match
www.celatro.com

Celatro™

Beat the competition with the fastest search algorithms available for .NET!

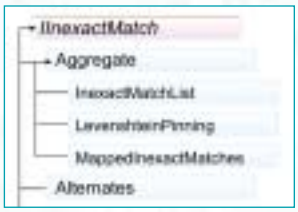
What is **pattern matching**?

In computer science, a *pattern* is defined as a *sequence of data elements*, and *pattern matching* as the process of *finding this sequence in data*. Any computer user who has used the **Find** function in a text editor has used a form of pattern matching known as *exact matching*. Similarly, anyone who has used a spell-checker or a wildcard search has used a form of *inexact matching*.

Pattern matching has a wide range of applications, from word processing to Internet surfing to forensic DNA analysis. As data stores grow to include larger, more varied, and less structured data, so does the need for faster, more accurate, and more efficient pattern-matching technologies.

Celatro™: a **pattern matching library** for .NET

Leveraging the latest advances in string theory and pattern matching, Celatro™ features a diverse library of efficient lexical and linguistic algorithms, all of which have been optimized for the .NET framework.



With its rich toolset and flexible framework, Celatro can improve the performance and capabilities of nearly any software application.

- Celatro features a diverse assortment of exact, inexact, and linguistic-based algorithms.
- Celatro's algorithms have been finely tuned and optimized to run under managed application frameworks.
- Celatro quickly ranks and orders results based on their similarity to a specified pattern.
- The Celatro core library can be easily extended with optional plug-ins, such as **Fuzzy Finder**™, **Russian Name Search**™, and **Arabic Name Search**™.

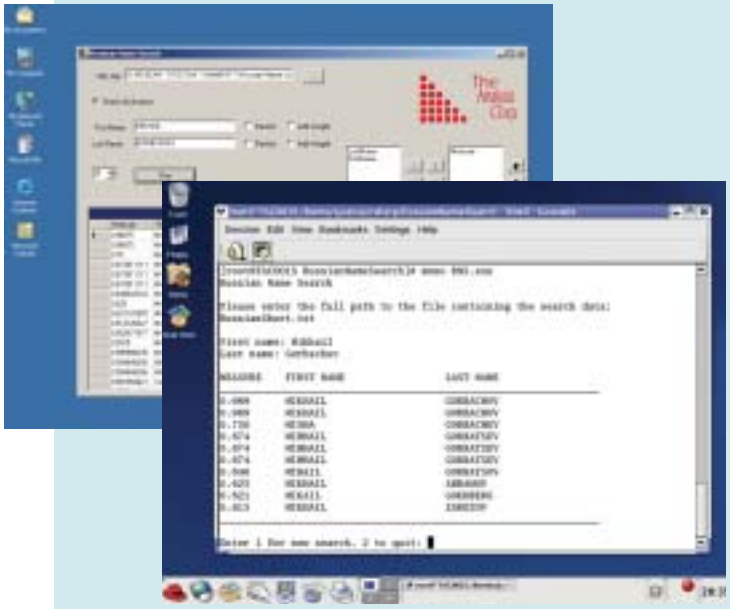
Highly **inter-connectible classes**

In addition to their ease of use and high performance, Celatro™ components are highly inter-connectible. Such extensibility allows developers to quickly devise sophisticated aggregates for specialized pattern matching needs.

Why .NET?

.NET offers:

- Cross-platform compatibility:
  - Celatro™ successfully executes on .NET ports such as Mono (see www.go-mono.com), and will run under Windows, Unix, Linux, and Mac OSX.
- Superior development platform:
  - .NET languages support object-oriented design idioms and good practices.
  - Visual Studio.NET lowers development costs and accelerates time-to-market.
  - The managed execution environment eliminates resource leaks and allows for run-time optimization.
- Microsoft's next-generation framework:
  - Future versions of Windows will require applications to be written in .NET.
  - .NET allows interoperability with deprecated technologies such as COM and MTS, as well as Microsoft products such as Outlook, Excel, and Word.
  - .NET assemblies have significant security advantages over executable binaries.
  - .NET provides native support for XML and SOAP.



Celatro™ is cross-platform compatible.

Celatro™ multilingual support

Celatro™ offers a wealth of features to address the difficulties of finding patterns in data written in different languages. Language-specific and customizable components, such as alphabets, tokenizers, and sentence delimiters, provide an easy means to parse both natural and custom languages. Furthermore, Celatro's exact and inexact matching algorithms have been tuned for large alphabets such as Unified CJK, and offer superior performance to native .NET string functions.

Celatro™ language-specific technologies for transliterated data

Finding proper names in transliterated data is difficult because of the way words are phonetically translated into English. Languages such as Russian and Arabic, which are written in scripts fundamentally different from English, yield transliterated data that is notoriously difficult to search through. For example, the common Russian equivalent of "Joseph" can be transliterated variously from Cyrillic script as Josif, Iosiff, Yosiph, or Jozeph.

Celatro™ enables developers to overcome these difficulties through the combined application of algorithmic and linguistic technologies. Additional plug-ins, such as Russian Name Search™ and Arabic Name Search™, extend Celatro capabilities with out-of-the-box solutions for specific languages.

Celatro™

Celatro™ plug-ins

In addition to its core pattern matching library, the Celatro™ family of technologies includes optional plug-ins that can be used to provide pattern matching capabilities for specific applications. Celatro plug-ins include:

- **Fuzzy Finder™:** This module handles *large spectrum data corpora*, such as those containing both alphabetic and numeric data elements. It provides a measure of similarity between each query and a *search corpus*, as well as a *pinning* mechanism that allows users to specify that certain characters or numbers must appear at specific locations or regions in the search data. The Celatro™ Fuzzy Finder plug-in extends our proven Fuzzy Finder technology with additional options and increased performance.
- **Russian Name Search™:** Russian Name Search is Celatro's pattern matching product specifically geared at retrieving different transliterations of Russian personal names. On top of its robust pattern matching mechanism, Russian-specific algorithms, such as nickname matching and spelling variant substitutions, increase both the precision and recall of Russian name retrieval.
- **Arabic Name Search™ :** Arabic Name Search is Celatro's pattern matching plug-in designed for the retrieval of Arabic personal names. Its sophisticated Arabic-specific parsing technologies identify name components such as prefixes, articles, roots, and suffixes, regardless of spelling variations. Arabic Name Search produces accurate results even when operating on names that are incomplete or have unexpected transliterations into English.

About us

Celatro™ is built upon fourteen years of successful algorithmic, linguistic, and criminological work by The Analysis Corp. (TAC) of Fairfax, Virginia. A partner of the federal government, TAC has developed products such as TIPOFF and Fuzzy Finder™, both of which have met with widespread success in the fields of crime prevention, fraud detection, and counter-terrorism.

Fuzzy Finder™ was designed to match inconsistent, incomplete, or erroneous alphanumeric data. Integrated into TIPOFF in 1990, Fuzzy Finder has been continuously operational at the DOS in support of the FBI, CIA, DIA, and NSA, and has become the most effective tool for finding the names of suspected or known terrorists in databases. Fuzzy Finder was extended with Arabic Name Search™ in 1997 to search terrorist watch lists provided by the CIA, NSA, and DIA. In 2000, it was further enhanced with Russian Name Search™ as part of the fight against Russian organized crime at the DOS.

The Analysis Corp. (TAC), renowned for developing specialized search systems for federal government intelligence agencies, provides high-technology computer systems engineering, directed research and development, and business management services to public and private clients.

www.theanalysiscorp.com

SFA's Systems Engineering Division, Applied Technology Division, and Frederick Manufacturing Division (FMD) provide high-technology services and prototyping to government and industry. Since 1969, SFA has capitalized on leading-edge systems and technologies and entered the intelligence and law enforcement arenas by acquiring The Analysis Corp. in 2003.

www.sfa.com

The Analysis Corp., 8613 Lee Hwy., Suite 250, Fairfax, VA 22031 ▪ 703-208-9630 ▪ Fax 703-208-9634