

SYSTRAN

Linguistic Development Kit (LDK)

Erstellen Sie Lösungen speziell für Ihre Kunden

Da die Entwicklung maschineller Übersetzungslösungen Spitzentechnologien erfordert, ist SYSTRAN seit Jahrzehnten bei der Verarbeitung linguistischer Daten und natürlicher Sprache führend tätig.

Zu unseren unterstützten Sprachen bieten wir eine Reihe wichtiger linguistischer Funktionen:

- Dokumentenfilterung
- Erkennung der Sprache
- Segmentierung und Tokenisierung
- Sprachnormalisierung
- Dokumentenklassifizierung (Fachgebietserkennung)
- Eigennamenerkennung
- Wörterbücher
- Morphologische Analyse
- Syntaktische Analyse
- Transliteration
- Wortbedeutungserklärung



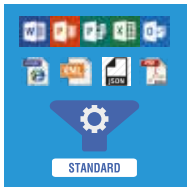
Tool zur Sprach-, Dokumenten- oder Namensverarbeitung. Durch eine Kombination können Sie von leistungsstarken, mehrsprachigen Lösungen für Data-Mining- oder semantische Suchsoftwareprogramme profitieren.

Einzeln genommen ist jedes Modul ein effizientes. Unabhängig von den Anforderungen Ihrer Endkunden (E-Discovery und digitale Forensik, OSINT- oder COMINT-Analysen, Wettbewerbs- und Marketinginformationen, Überwachung der Online-reputation, Stimmungsanalyse für mehr Kundenverständnis) ermöglicht Ihnen das SYSTRAN LDK die Analyse strukturierter und unstrukturierter mehrsprachiger Inhalte wie von Benutzern erzeugte Inhalte, soziale Medien, Web-Inhalte und vieles mehr.

Jetzt verfügen Sie über die komplette Ausstattung zur Erzeugung leistungsstarker Data-Mining- oder Intelligence-Lösungen für Ihre innovationsfreudigen Kunden!

Diese linguistischen Funktionen werden als **Software Development Kit** mit einem geringen Speicherbedarf bereitgestellt, sodass externe Entwickler und Technikpartner sie problemlos in ihre eigenen Anwendungen und Dienste integrieren können.

ANFORDERUNGEN	Native C++-Bibliotheken mit geringem Speicherbedarf und Bindungen für die gängigsten Programmiersprachen (Java, .NET, Javascript etc.)
PLATTFORM-SUPPORT	Windows, Linux, Mac OS, iOS, Android



Dokumentenfilterung

Importiert verschiedene Dokumentenformate zur Verarbeitung durch andere Module und Neuaufbau des Dokuments im Originalformat mit geändertem oder kommentiertem Inhalt und beibehaltenem Layout



Erkennung der Sprache

Erkennt automatisch die Dokumentensprache durch spezielle Wort- oder Satzmustererkennung



Segmentierung und Tokenisierung

Segmentiert den Text in Sätze und Sätze in Tokens (kleinste Verarbeitungseinheit)



Sprachnormalisierung

Normalisiert Text aus Blogs, E-Mails, Chatforen und von Benutzern erzeugten Inhalten und korrigiert dabei gängige Fehler und Sprachabweichungen



Dokumentenklassifizierung

Bestimmt das Fachgebiet des Dokuments auf Grundlage vordefinierter Modelle und zeigt die wichtigsten Schlüsselwörter an



Eigennamenerkennung

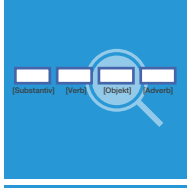
Erkennt auf Grundlage der Analyse der Dokumenteninhalte automatisch Personennamen, Orte, Zahlen, Daten und Organisationsnamen und zeigt diese an



Wörterbuch

Wörterbuch

Bietet ein- und zweisprachige Suchen mit zusätzlichen kontextuellen Informationen wie Bedeutung, Fachgebiet, Kontext, Ausdrücke und Beispiele



Morphologische Analyse

Bietet morphologische Analysen einzelner Wörter einschließlich einer Liste möglicher Lemmata und Wortarten für eine Flexionsform



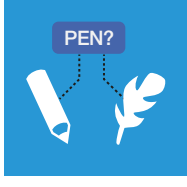
Syntaktische Analyse

Bietet syntaktische Analysen auf Satzebene mit folgenden linguistischen Analysen: Worterkennung, Part-of-Speech-Tagging sowie Bestandteil- und Abhängigkeitsanalysen



Transliteration

Erstellt Transkriptionen und Transliterationen von Wörtern oder Einheiten zwischen Sprachen mit verschiedenen Schriften und erkennt Eigennamenursprünge



Wortbedeutungserklärung

Wählt die beste Bedeutung eines Wortes in Abhängigkeit vom Kontext