Bitext NLP Pipeline Overview

Multiplatform & Enterprise-grade NLP SDK

The Bitext NLP Pipeline SDK supports a wide range of linguistic tasks including Language Identification, Sentence Segmentation, Tokenization, Lemmatization, Decompounding, POS Tagging, Entity Extraction and more. It works in 15 languages today (Q1 2026), including English, Spanish, Arabic, Chinese, Japanese and more.

The Bitext SDK is used by top industry players, including NASDAQ-listed companies. It has been engineered in C to achieve the top enterprise requirements

Main Multilingual CPU-Based NLP SDK Features

Multiplatform Design

Coded in platform-independent C. Operating Systems: Windows, Linux, macOS. Architectures: x64, ARM.

Simple Integration

The SDK integrates seamlessly via C, Python and Java bindings for maximum compatibility across development environments.

Deployment Flexibility

Supports both on-prem and cloud installations, via both REST API and on-premise API/SDK configurations.

Lightweight Architecture

The full pipeline uses 100MB storage footprint with no additional dependencies and 200MB memory usage.

High Compression

Compression rates of 1:100 (100MB of raw text data is compressed to 1MB) resulting in ultra-fast access: 400 million lookups per second in one 8-core CPU.

Maximum Performance

Processes over 640,000 words per second in one CPU (8-core) with ultra-low-latency processing

High Accuracy:

State-of-the-art (SOTA) F1-Score accuracy, evaluated using standard public benchmark corpora like PUD, CoNLL or WikiNEuRal

Customizable by design:

Our NLP engine can be tailored to your needs we easily add new entity types, custom taxonomies, or languages upon request. Whether you need domain-specific recognition or multilingual expansion, Bitext adapts to your business context without retraining or extra labeling.

640K+

Words per Second for Entity Extraction

High-performance processing on single 8-core CPU

100M

Queries per Second for Lemmatization

High-performance processing on single 8-core CPU

25M

Queries per Second for Decompounding

High-performance processing on single 8-core CPU

50MB

Storage Footprint

Per language with no additional dependencies

200MB

Memory Usage

Efficient resource utilization per language pipeline

1:100

Compression rate

External & Internal Data Compression Rate

No External Dependencies for Maximum Privacy Compliance

- Independent of external cloud architectures for maximum privacy
- No need for external data, like pre-tagged training data
- Self-contained software pack, with no external dependencies

Contact Us

Info@bitext.com

Learn More

www.bitext.com