

# Entity Benchmarking Report

The following report compares Bitext's entity extraction and 3 other engines (CRFSuite, Stanford and SENNA) on a well-known 2013 dataset, OntoNotes 5.

OntoNotes 5 is the final release of the OntoNotes project, a collaborative effort between BBN Technologies, the University of Colorado, the University of Pennsylvania and the University of Southern Californias Information Sciences Institute. The goal of the project was to annotate a large corpus comprising various genres of text (news, conversational telephone speech, weblogs, usenet newsgroups, broadcast, talk shows) in three languages (English, Chinese, and Arabic).

For this benchmark we have used the "Entity" tagging of OntoNotes 5, and specifically the "Weblogs" subset (around 5,500 named entities).

We have analyzed the subset with Bitext's Entity Extraction Service, obtaining entities and types from three different entity types: Persons, Places and Organizations/Companies.

We have compared Bitext's results with the results appearing in this 2017 paper: <http://www.sciencedirect.com/science/article/pii/S088523081630002X>, in which some independent researchers analyzed this corpus (and others) against 3 Entity Extraction engines: CRFSuite, Stanford and SENNA.

---

## Entity Extraction result comparison

### Entity & type matching

	Precision	Recall	F-Score
<b>Bitext</b>	77%	55%	64%
<b>CRFSuite</b>	65%	45%	53%
<b>Stanford</b>	69%	54%	60%
<b>SENNA</b>	64%	61%	62%

### Just entity matching (without type)

	Precision	Recall	F-Score
<b>Bitext</b>	89%	70%	78%

## Benchmarking procedure

### Evaluated Software

**For the benchmarking of Bitext we have used their API.**

<https://api.bitext.com/>

Taking as entity the “Entity” field from their Entity Extraction service.

The type has been taken from the “Type” field, with this correspondence:

“1” = “Person”

“3” = “Place”

“6” and “7” = “Organization/Company” (for comparison purposes we have merged Bitext’s “6” (*company*) and “7” (*organization*) types into just one (“*Organization/Company*”).

The results of CRFSuite, Stanford and SENNA have been extracted from the cited paper (<http://www.sciencedirect.com/science/article/pii/S088523081630002X>). The followed procedures are detailed there.

## Comparison metrics

We have used three widely used comparison metrics. “*Correct*” in this case means “*same entity and type*”.

**1. Precision** =  $\text{Correct entities} / (\text{Correct entities} + \text{Incorrect entities})$

**2. Recall** =  $\text{Correct entities} / (\text{Correct entities} + \text{Missed entities})$

**3. F Score** =  $2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}))$