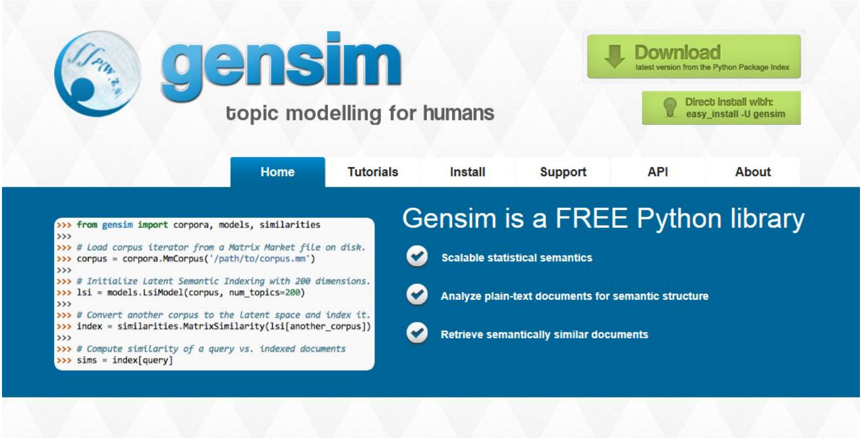
Text
Mining
Analysis
Analytics
Application
Information
Research
Data
Biomedical
Entity
Involve
Business
Word
Pattern
Such
Set
New
Software
Search
Document
Literature
Extract
Sentiment
Describe
Large
Online
Name
Web
Bi
Process
Need
Content
Social
Media
Intelligent
Edit
Use

gensim

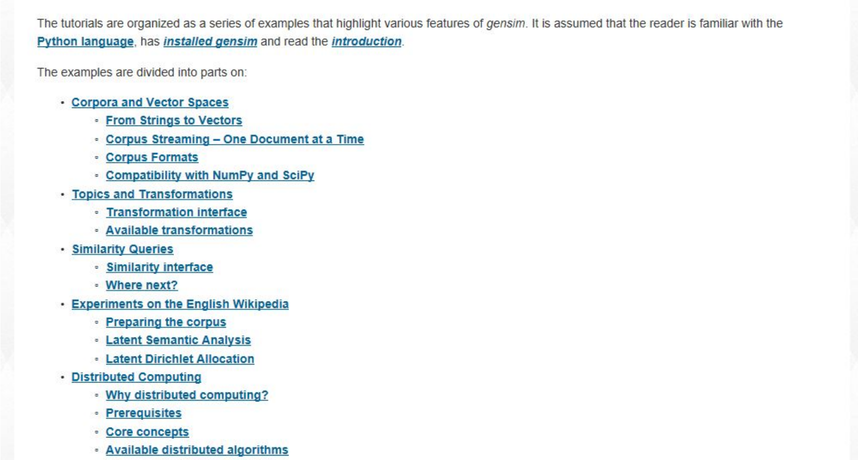


```
>>> from gensim import corpora, models, similarities
>>>
>>> # Load corpus iterator from a Matrix Market file on disk.
>>> corpus = corpora.MmCorpus('/path/to/corpus.mm')
>>>
>>> # Initialize Latent Semantic Indexing with 200 dimensions.
>>> lsi = models.LsiModel(corpus, num_topics=200)
>>>
>>> # Convert another corpus to the latent space and index it.
>>> index = similarities.MatrixSimilarity(lsi[another_corpus])
>>>
>>> # Compute similarity of a query vs. indexed documents
>>> sims = index[query]
```

Gensim is a FREE Python library

- ✓ Scalable statistical semantics
- ✓ Analyze plain-text documents for semantic structure
- ✓ Retrieve semantically similar documents

gensim

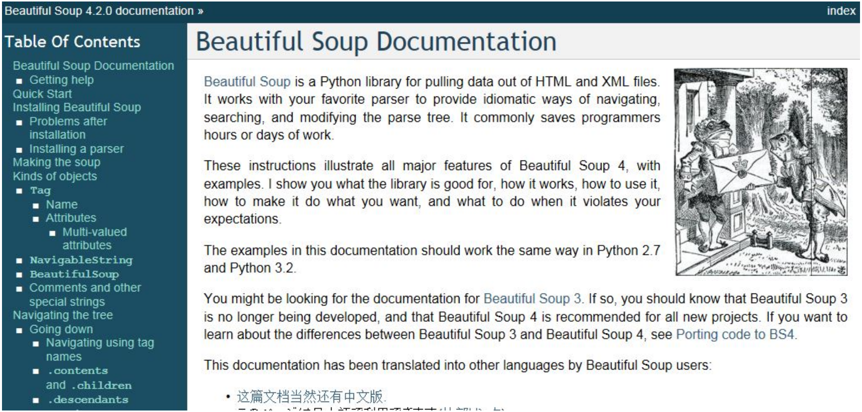


The tutorials are organized as a series of examples that highlight various features of *gensim*. It is assumed that the reader is familiar with the **Python language**, has ***installed gensim*** and read the ***introduction***.

The examples are divided into parts on:

- Corpora and Vector Spaces
  - From Strings to Vectors
  - Corpus Streaming – One Document at a Time
  - Corpus Formats
  - Compatibility with NumPy and SciPy
- Topics and Transformations
  - Transformation interface
  - Available transformations
- Similarity Queries
  - Similarity interface
  - Where next?
- Experiments on the English Wikipedia
  - Preparing the corpus
  - Latent Semantic Analysis
  - Latent Dirichlet Allocation
- Distributed Computing
  - Why distributed computing?
  - Prerequisites
  - Core concepts
  - Available distributed algorithms

Beautiful Soup



Beautiful Soup 4.2.0 documentation »

Table Of Contents

Beautiful Soup Documentation

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

These instructions illustrate all major features of Beautiful Soup 4, with examples. I show you what the library is good for, how it works, how to use it, how to make it do what you want, and what to do when it violates your expectations.

The examples in this documentation should work the same way in Python 2.7 and Python 3.2.

You might be looking for the documentation for Beautiful Soup 3. If so, you should know that Beautiful Soup 3 is no longer being developed, and that Beautiful Soup 4 is recommended for all new projects. If you want to learn about the differences between Beautiful Soup 3 and Beautiful Soup 4, see Porting code to BS4.

This documentation has been translated into other languages by Beautiful Soup users:

- 这篇文档当然还有中文版.

Beautiful Soup

Beautiful Soup

- Beautiful Soup is a Python library for pulling data out of HTML and XML files.
- It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree.
- It commonly saves programmers hours or days of work.