



链滴

# Java 自然语言处理相关的资料

作者: [felayman](#)

原文链接: <https://ld246.com/article/1519819192856>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fwww.cortical.io%2F" target="\_blank" rel="nofollow ugc">Cortical.io</a> – 像人脑一样快速、精确处理复杂的 NLP（自然语言处理）操作（如消歧、分类、流文本过滤等操作）的 Retina API； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fcorenlp.shtml" target="\_blank" rel="nofollow ugc">CoreNLP</a> – 斯坦福大学的 CoreNL 提供的一系列的自然语言处理工具，该工具可以根据输入原始英语文本，给出单词的基本形式； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fflex-parser.shtml" target="\_blank" rel="nofollow ugc">Stanford Parser</a> – 一种自然语言分析器，可以分析语句的语法结构； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Ftagger.shtml" target="\_blank" rel="nofollow ugc">Stanford POS Tagger</a> – 一个词性类器 (POS Tagger) ； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2FCRF-NER.shtml" target="\_blank" rel="nofollow ugc">Stanford Name Entity Recognizer</a> – Stanford NER 是一个 Java 实现的名词识别器； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fsegmenter.shtml" target="\_blank" rel="nofollow ugc">Stanford Word Segmenter</a> – 分词器，很多 NLP 工作中都要用到的标准预处理步骤； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Ftregex.shtml" target="\_blank" rel="nofollow ugc">Tregex, Tsurgeon and Sengrex</a> – regex 基于树关系以及节点匹配的正则表达式，用于在树状数据结构中进行模式匹配（名字是 “tree regular expressions” 的缩写） ； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fphrasal%2F" target="\_blank" rel="nofollow ugc">Stanford Phrasal: </a>一个基于短语的译系统</li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Ftokenizer.shtml" target="\_blank" rel="nofollow ugc">Stanford English Tokenizer</a> – Stanford Phrasal 用 Java 写成的最新的基于统计短语的机器翻译系统； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Ftokensregex.shtml" target="\_blank" rel="nofollow ugc">Stanford Tokens Regex</a> – 一个分解器，可以将文本大致分成一系列对应于 “词” 的符号； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fsutime.shtml" target="\_blank" rel="nofollow ugc">Stanford Temporal Tagger</a> – SUTme 是一个用于识别并标准化时间表达式的库； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Fpatternslearning.shtml" target="\_blank" rel="nofollow ugc">Stanford SPIED</a> – 在种集上使用模式，以迭代方式从无标签文本中进行学习实体； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fnlp.stanford.edu%2Fsoftware%2Ftmt%2Ftmt-0.4%2F" target="\_blank" rel="nofollow ugc">Stanford Topic Modeling Toolbx</a> – 为社会科学家及其他希望分析数据集的人员提供的主题建模工具； </li>
- <li> <a href="https://ld246.com/forward?goto=https%3A%2F%2Fgithub.com%2Ftwitter%2Ftwitter-text-java" target="\_blank" rel="nofollow ugc">Twitter Text Java</a> – Java 实现的推特文本处理库； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Fmallet.cs.umass.edu%2F" target="\_blank" rel="nofollow ugc">MALLET</a> – 基于 Java 的统计自然语言处理、文档分类、聚类、主题建模、信息提取以及其他机器学习的文本应用包； </li>
- <li> <a href="https://ld246.com/forward?goto=https%3A%2F%2Fopennlp.apache.org%2F" target="\_blank" rel="nofollow ugc">OpenNLP</a> – 基于机器学习的自然语言文本处理工具包； </li>
- <li> <a href="https://ld246.com/forward?goto=http%3A%2F%2Falias-i.com%2Flingpipe%2Findex.html" target="\_blank" rel="nofollow ugc">LingPipe</a> – 一个使用计算机语言学文本处理工具包； </li>
- <li> <a href="https://ld246.com/forward?goto=https%3A%2F%2Fcode.google.com%2Fp%2Fc" target="\_blank" rel="nofollow ugc">Stanford CoreNLP</a> – 斯坦福大学 CoreNLP 的源代码； </li>

earTk%2F" target="\_blank" rel="nofollow ugc">ClearTK</a> – 在 Java 中 ClearTK 为开发统计  
言处理组件提供了一个框架，该框架是基于 Apache UIMA; </li>  
<li><a href="https://ld246.com/forward?goto=http%3A%2F%2Fctakes.apache.org%2F" target  
="\_blank" rel="nofollow ugc">Apache cTAKES</a> – Apache cTAKES 是一个开源自然语言处  
系统，用于从临床电子病历的自由文本中提取信息; </li>  
<li><a href="https://ld246.com/forward?goto=http%3A%2F%2Fwww.clearnlp.com%2F" targ  
t="\_blank" rel="nofollow ugc">ClearNLP</a> – ClearNLP 工程为自然语言处理提供了软件和资  
提供了。该项目最早在计算机愈合和教育研究中心启动，目前由 Emory 大学的语言和信息研究中心  
续开发。该项目遵循 Apache 2 license。 </li>  
</ul>