

# Semantics API 及演示样例 介绍

Ver 0.1.1218

## 1 总体简介

Semantics API 是一套 NLP 处理云平台。先期推出语义句法分析云服务。

Semantics API 采用基于 HTTP AJAX/JSON 接口的 WebAPI 技术提供开放的服务接口。为了便于开发，Semantics API 提供了 Java, JavaScript 和 .NET 的 SDK 包。

后续计划提供 Python SDK。在 Python SDK 发布之前，可以考虑利用 IronPython 技术调用 .NET SDK。

### 1.1 可学习的 NLP 系统

鉴于应用的千差万别，目前技术水平的 NLP 系统一般都不能覆盖所有语言现象。包括词义词用，或者句法表达方式。因而 Semantics API 设计成一种可学习，可扩展的架构。扩展性体现在：

- 用户词典管理：
- 用户语料库管理：
- 歧义检测报告与指正：特别是针对多义词词义歧义(Word Sense Ambiguity, WSA)提供了机制来教学指正。

Semantics API 提供接口的接口包括：

- 歧义报告接口：在 Parse 返回结果中，定义了数据结构来描述所遇到的歧义
- 歧义指正接口：提供函数调用将客户指正的消歧知识提交给 Semantics 系统，在后续 Parse 中加以应用。

## 2 Java API 介绍

### 2.1 编译运行

- 1) 用 Eclipse （推荐用 4.7Oxygen 版本）从 <https://github.com/graphysics/semantics.git> 获取所有源程序，
- 2) 更新 maven，下载 codehaus, jackson, log4j 等组件
- 3) 运行 SwingApplication.main。

### 2.2 安全机制：回话握手与登录

Semantics WebAPI 云服务基于互联网基础设施。为了保证安全性，设计提供基于回话令牌的安全机制。

安全机制的另一目标是保持不同客户应用之间的数据安全隔离。

应用启动时应当调用 ISemantics.Hello 函数来与 WebAPI 服务握手，得到 Aession 类型数据，并缓存。后续用于维护令牌等缓存对象。

ISemantics.Login 执行登陆操作，返回 Aoken 类型的令牌对象。客户应用应当保存此对象在 Aession 之中，后续的 API 函数调用会自动从 Aession 中读取 Token，作为默认参数，和其他接口函数参数一起传递给 WebAPI 服务。

（说明：当前演示版本的安全机制并不完善，而在于打架安全框架与模型。未来 Release 会逐步完善。）

### 2.3 Parsing

ISemanticParser 提供 Parse 等函数调用。

Parse 调用返回 ParseCase 对象，包括成功的句法分析结果或错误原因、歧异等。Parse 中会自动检测不同通过的错误并报告。

下图演示了 Semantics Parser 返回的 Parse 结果。其中右侧为 WebAPI AJAX 调用返回的 JSON 原始报文文本。



ParseCase.Result 是一个 Aode 类型对象(API Node)。Aode 本质上是一个树数据结构的节点，Aode.Children 成员记录一个树节点的所有下级节点，层次递进知道词语叶节点。

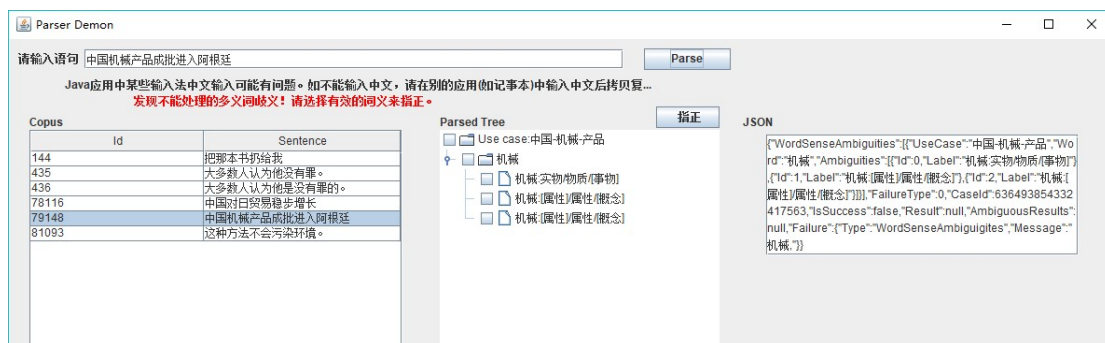
## 2.4 指正消歧

Parse 不能成功的原因往往是因为歧义。Semantics 应用上下文消歧等多种消歧技术。但需要客户提供消歧知识规则。Java 演示应用中，演示了如何调用 IsemanticParser 接口中的 Tutor 函数辅助消歧。

### 2.4.1 词义指正消歧

Semantics API 在句法分析同时会对多义词执行词义辨析，以保证句法分析准确性。词义消歧基于词义共现等知识，成为词用知识。如果客户提交的语句中的词语在 Semantics API 系统中没有词用知识，那么需要客户对词义进行辨析指正，系统会记录指正的词用，在后续句法分析中加以应用。

如下图所示为一个词义消歧指正实例：



如上图所示的短语‘中国-机械-产品’中，系统因缺乏词用共现知识，不能理解‘机械’一词在这里的含义。既可能是指‘机械/教条/僵化’这样的形容词，也可能是指‘机械/机器’这样的名词。

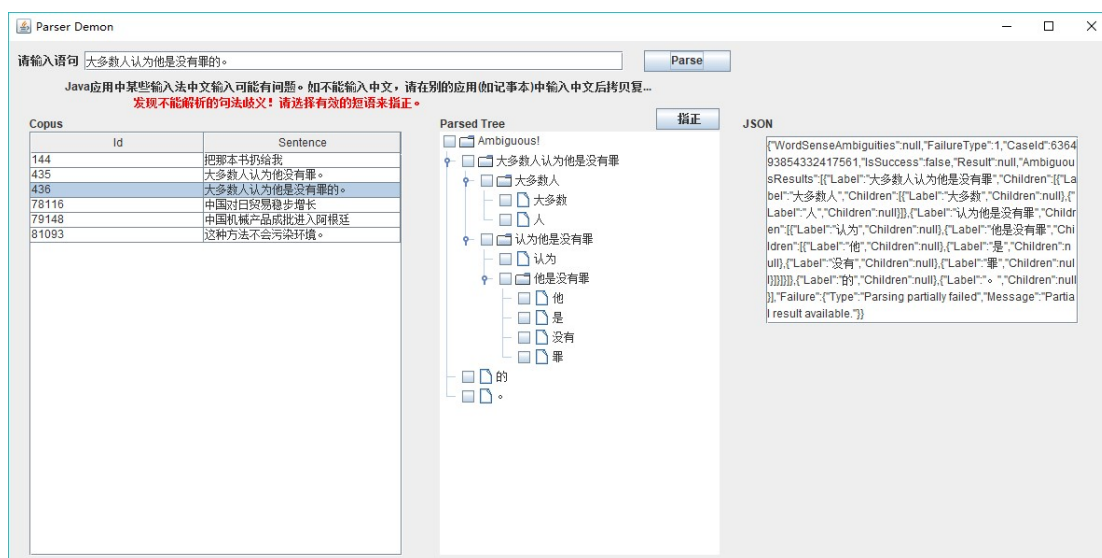
当出现词义歧义时，Parser Demo 窗体中的 ParseTree 控件显示的将是词义歧义情况。根节点为短语词用，1 级子节点为歧义的词语。2 级叶节点则是该词语

的若干义项。这时用户需要选择‘机械/实物/物质/[事物]’词义项,然后点击‘指正’按钮提交。

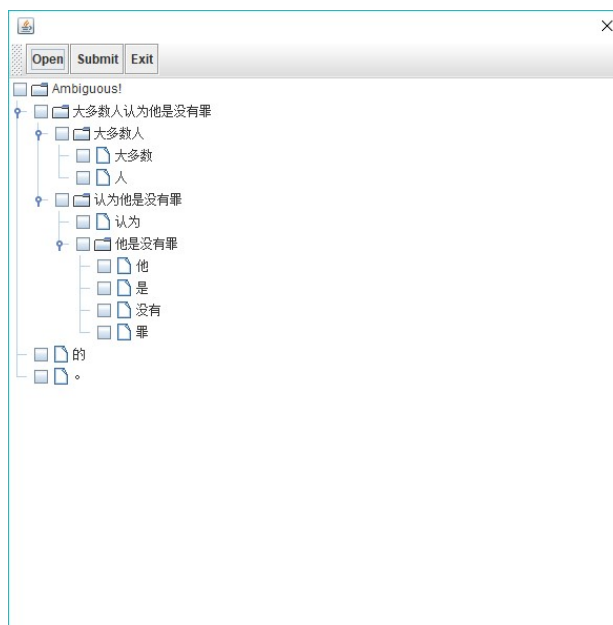
一个短语中可能有多个词语存在歧义。

### 2.4.2 句法指正消歧

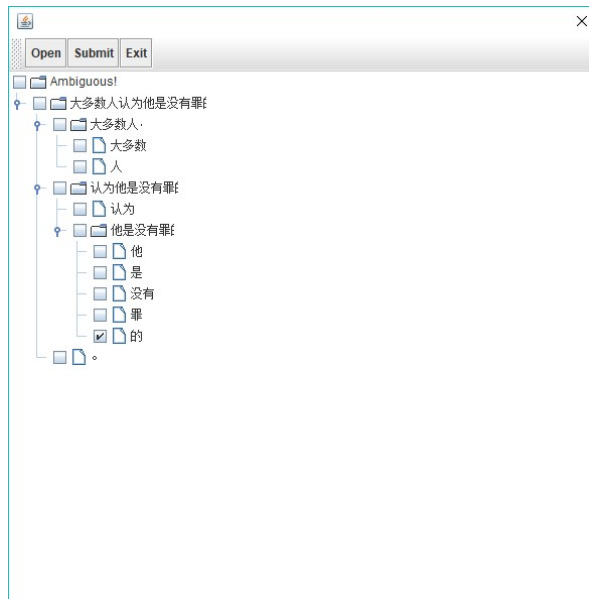
当句法分析失败时，**Semantics API** 会返回部分分析的结果。如下图用例中，**Semantics NLP** 不能成功完成整句的理解分析，而得到三段结果。并且第一段结果还有错误。



此时可以点击“指正”按钮进入指正操作界面:



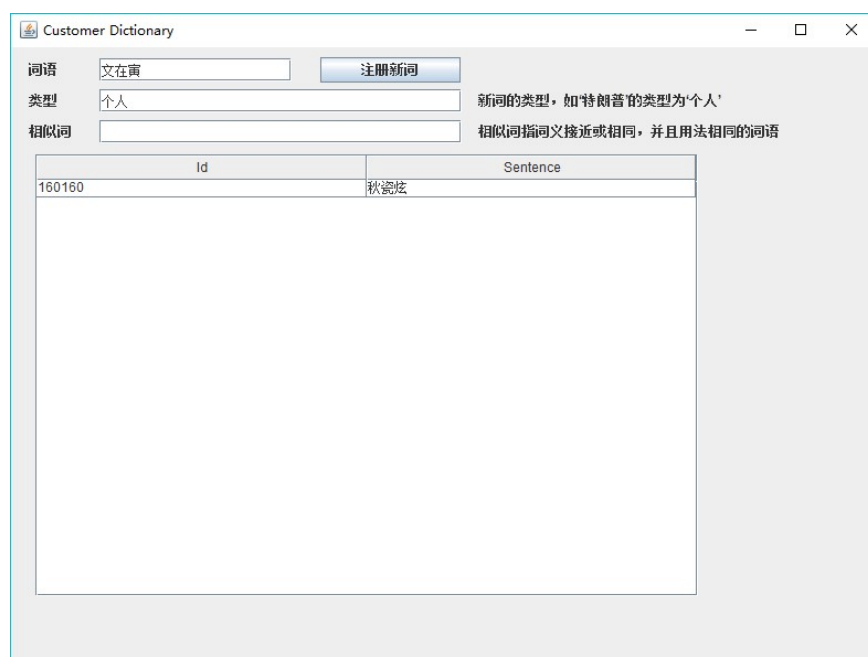
通过拖放树节点操作，可以得到正确的句法树：



此时可以点击“Submit”提交到 Semantics 平台。系统会记录此种正确分析。未来遇到类似的短语会给出相似的句法分析。

## 2.5 用户 Corpus

## 2.6 用户词典管理



### 3 JavaScript API 介绍

Semantics API 基于 WebAPI 技术，可以直接从浏览器或 node.js 等 javascript 应用中调用。演示样例也提供了一个 JavaScript 版本。但 JavaScript 版本仅演示了基本的登陆和 Parse。更多全面功能演示请参见.NET 或 Java 演示应用。



示例中定义了 form

```
<form id="form1" method="post">
  <input id="__token__" name="__token__" type="hidden" />
  输入语句: <input id="words" name="words" type="text" value="中国的首都" style="width:300px;" />
  <input id="btnParse" type="button" value="Parse" onclick="doparse()" style="cursor:pointer;" />
</form>
```

其中

```
function doparse() {
    $("#errorMessage").innerHTML = "";
    var txtbox = $("#words")[0];
    var postdata = $("#form1").serialize();
    var serviceurl = $("#serviceurl")[0].value + 'Parse';
    $.ajax({
        type: 'post',
        url: serviceurl,
        data: postdata,
        cache: false,
        success: function (data) {
            showparseresult(data);
            var divjson = $("#parseresult")[0].innerHTML = JSON.stringify(data);
        },
        error: function (e) {
            var div = $("#errorMessage")[0];
            if (e.responseText == null)
                div.innerHTML = e.statusText;
            else {
                var response = e.responseText;
                var failure = eval("(" + response + ")");
                div.innerText = failure.Type + ':' + failure.Message;
            }
        }
    });
}
```

```
}  
    }  
};  
}
```

## 4 问题处理

开发阶段，Semantics NLP 云服务不一定问题，可能会出现服务掉线的情况。如果遇到问题可以联系 Semantics NLP 云服务开发团队。