

# 5 brat 使用手册

原网页：<http://brat.nlplab.org/manual.html>

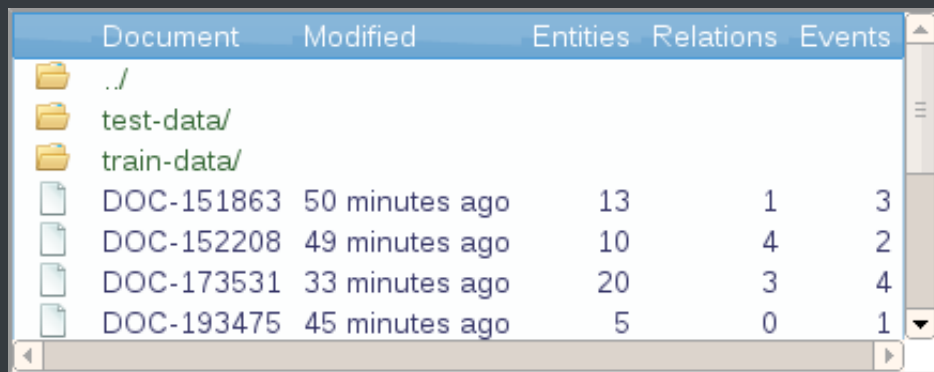
翻译：eillsu

GitHub：<https://github.com/eillsu/Brat-Chinese-Translation>

本文档介绍了 brat 快速标注工具的基本功能。

已熟悉 brat 但是该工具有问题？请参阅故障排除（<http://brat.nlplab.org/troubleshooting.html>）。

## 选择文档



Document	Modified	Entities	Relations	Events
../				
test-data/				
train-data/				
DOC-151863	50 minutes ago	13	1	3
DOC-152208	49 minutes ago	10	4	2
DOC-173531	33 minutes ago	20	3	4
DOC-193475	45 minutes ago	5	0	1

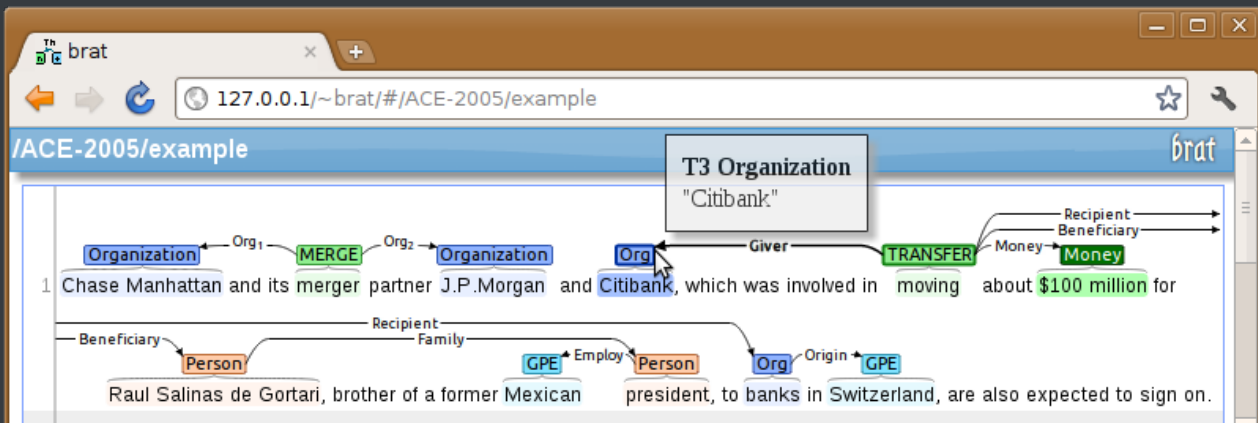
### 文档选择

启动时，brat将显示集合浏览器。要打开文档，只需双击文档选择器中的一个文档📄即可。

打开另外一个文档集合，双击集合📁。然后就会显示该集合的内容。

## 可视化

选择文档后，主窗口区域将显示该文档的文本和标注的可视化。

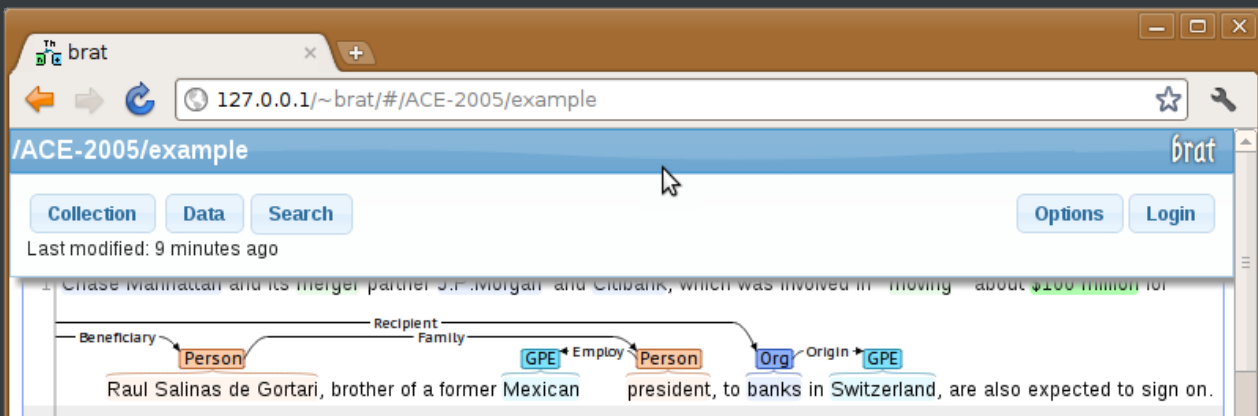


标注可视化

将鼠标光标放在标注上会显示有关该标注的一些进一步的信息，包括标注类型的完整形式、标注ID以及附加到标注的任何属性或标注。

## 菜单

将鼠标光标放在窗口的顶部栏上会打开菜单，该菜单从屏幕顶部向下滑动。



brat 菜单

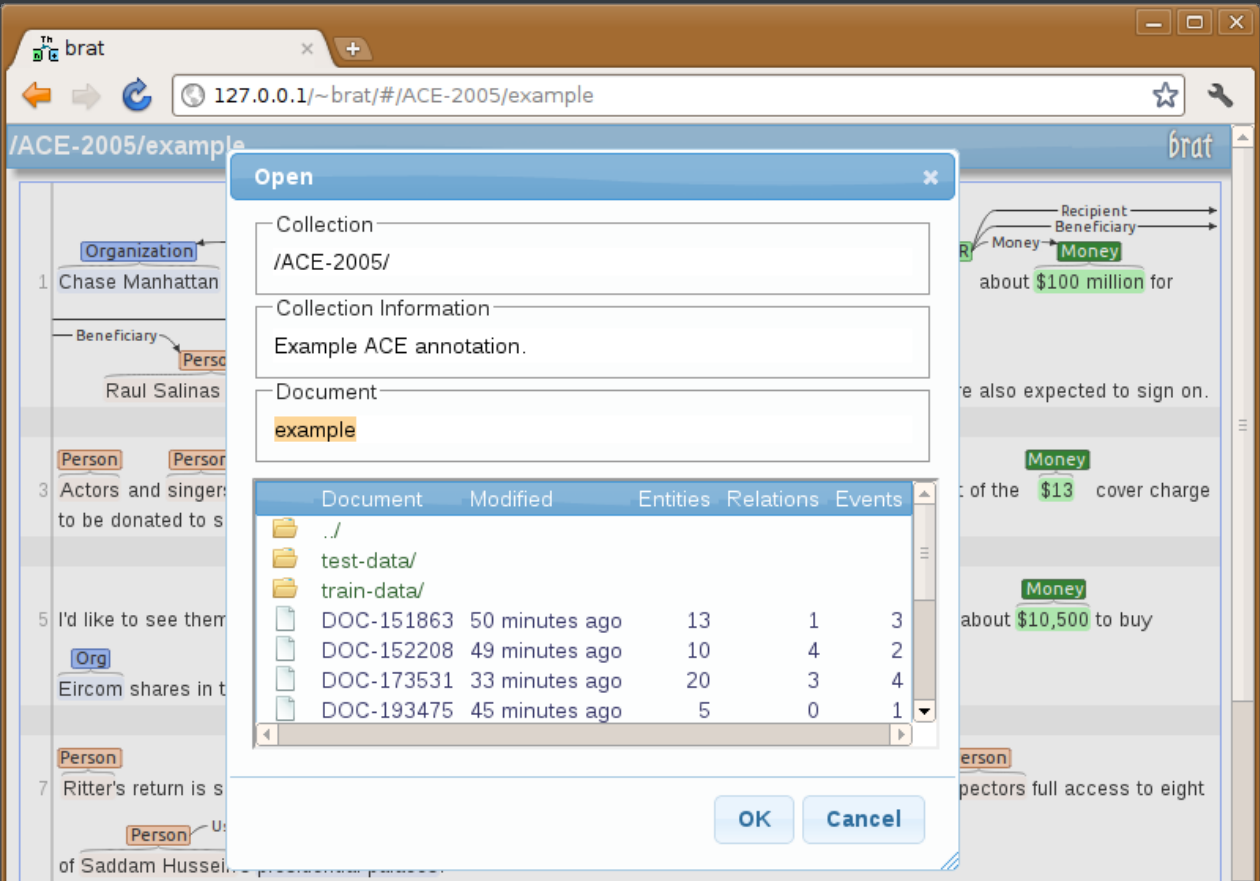
一旦鼠标指针离开菜单区域，菜单就会消失，从而通过仅在需要时出现来为标注提供更多的屏幕空间。

该菜单提供对以下功能的访问：

- **Collection:** [collection browser](#) 集合浏览器
- **Data:** [export/import/modify annotation data](#) 导出、导入、修改标注数据
- **Search:** [search current document or collection](#) 检索当前文档或者文件夹
- **Options:** system configuration options 系统配置选项
- **Login:** [log in for editing](#) 登录界面

## 集合浏览器

可以从菜单或按TAB键访问的集合浏览器，允许您访问安装中设置的那些集合中的不同文本集和单个文档。



### brat 集合浏览器

集合可以按层次结构组织：集合可以包含其他集合，类似于磁盘上的目录或文件夹结构。

其他集合在浏览器的前面显示。条目“../”指的是包含当前一个（如果有的话）的集合（父集合）。

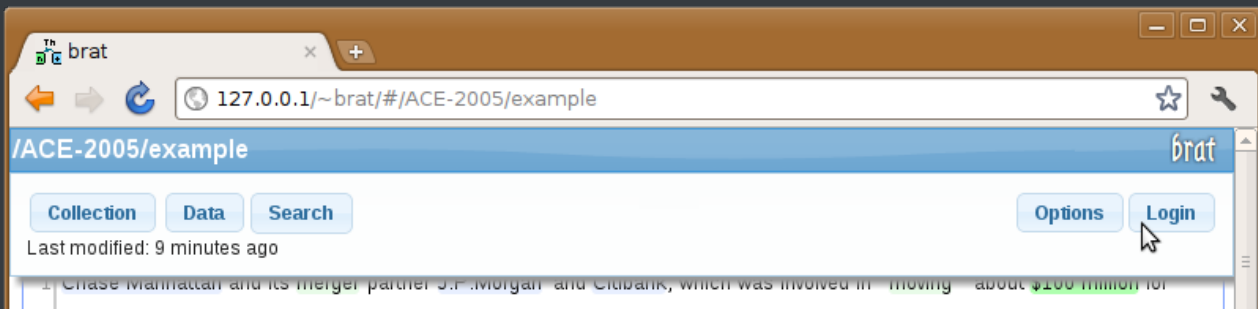
浏览器前面显示的是集合（文件夹），后面显示的是文档。对于每个文档，显示最后一次修改时间和简单标注统计信息（实体数、关系数甚至标注数）。

默认情况下，文档按文档名称排序。通过单击浏览器标题中的标签，可以按照其他条件对集合浏览器中的文档进行排序。例如，在“实体”上单击一次将对文档进行排序，以便首先显示具有最大数量的实体标注的文档。再次单击同一标签将反转此排序顺序，以便首先显示具有最少实体标注的文档。

❗ 以点（“.”）开头的名称将不会在集合浏览器中列出。此功能可用于创建“隐藏”集合。

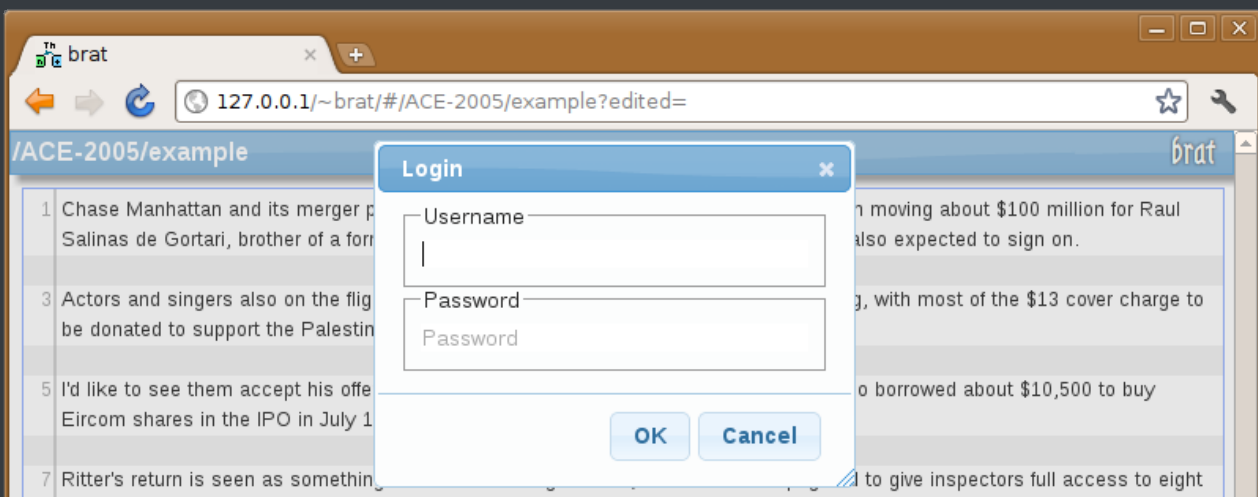
## 登录

要登录系统，首先单击菜单中的“登录”按钮。



登录按钮

单击“登录”按钮将显示登录对话框。



登录对话框

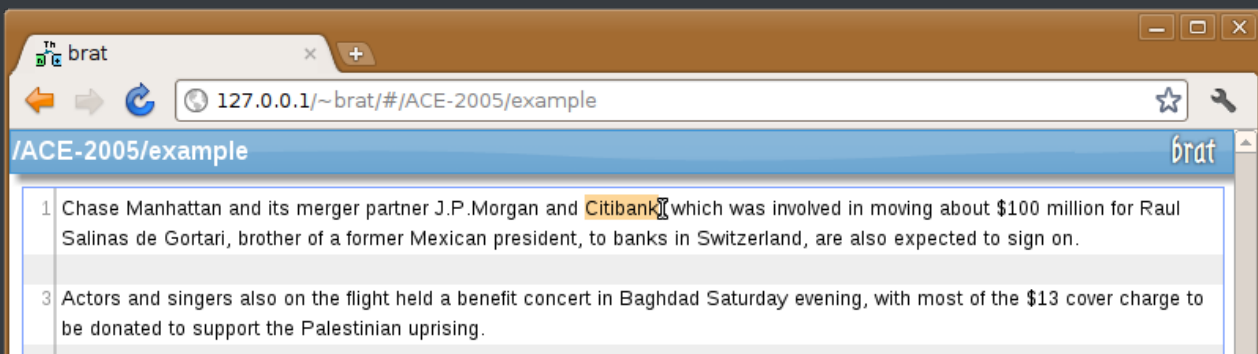
要登录，请在登录对话框中输入有效的用户名和密码。在服务器中设置用户和密码。如果您需要新的登录或忘记密码，请联系您的brat 管理员。

## 基本编辑

要编辑标注，用户首先需要使用具有编辑权限的帐户登录。

### 添加文本范围标注

登录后，只需使用鼠标选择该范围，即可为文本范围添加标注。



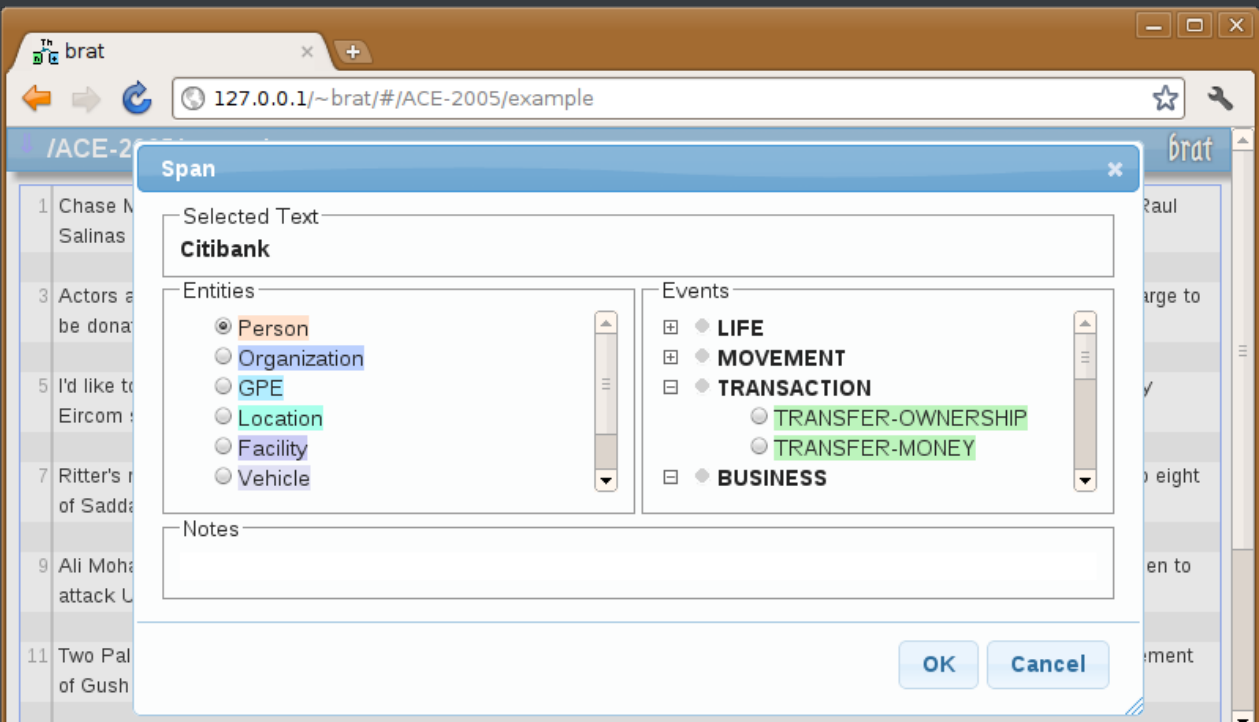
选择标注的文本

可以通过“拖动”文本或双击单词来进行选择。

❗ 也可以通过单击跨中的第一个和最后一个词来选择多词范围，如下所示：首先，按住ctrl，同时双击其中一个词，然后释放ctrl并按住shift，同时单击另一个词。

❗ 空范围，即文本中不包含任何字符的点，可以在单击点时按住SHIFT键来选择。

选择跨度后，系统将显示“跨度标注”对话框。



跨度标注对话框

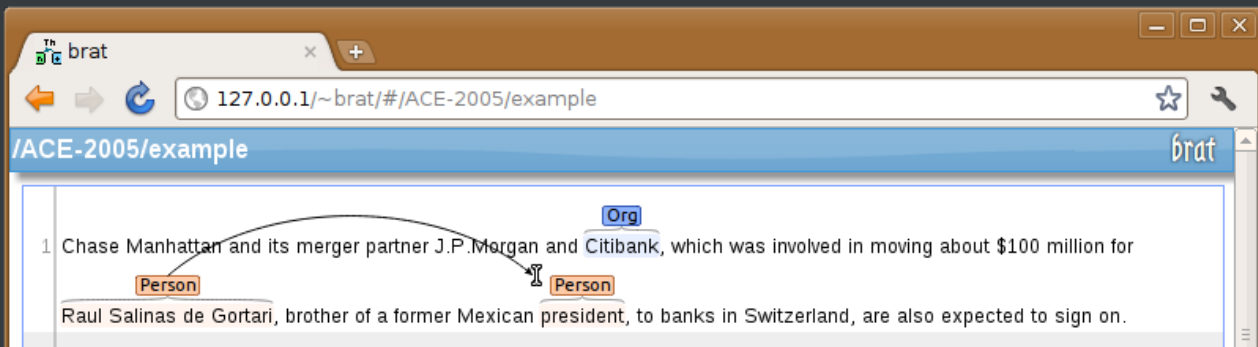
允许选择要分配给新创建标注的类型，以及添加标注或设置标注的其他方面，例如属性值（如果可用）。

⚠ 如果在选择文本时无法打开批注对话框，请首先检查您是否登录到具有编辑权限的帐户。失败的另一个可能原因可能是使用了不支持在SVG文本元素中选择的浏览器。请考虑使用支持的浏览器进行测试。

可用于标注的类型和可用于标注的其他方面取决于当前集合的配置。

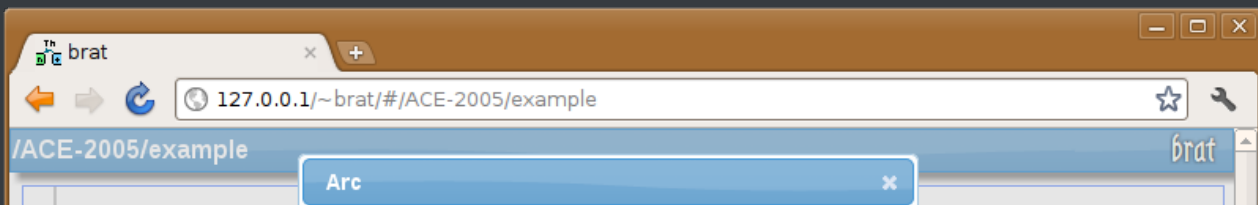
### 添加标注之间的关联

可以通过将鼠标从一个标注“拖动”到另一个标注来创建现有标注之间的关联。



关联已有标注

将鼠标指针重新释放到另一个标注上会再次显示一个用于选择标注类型的对话框。



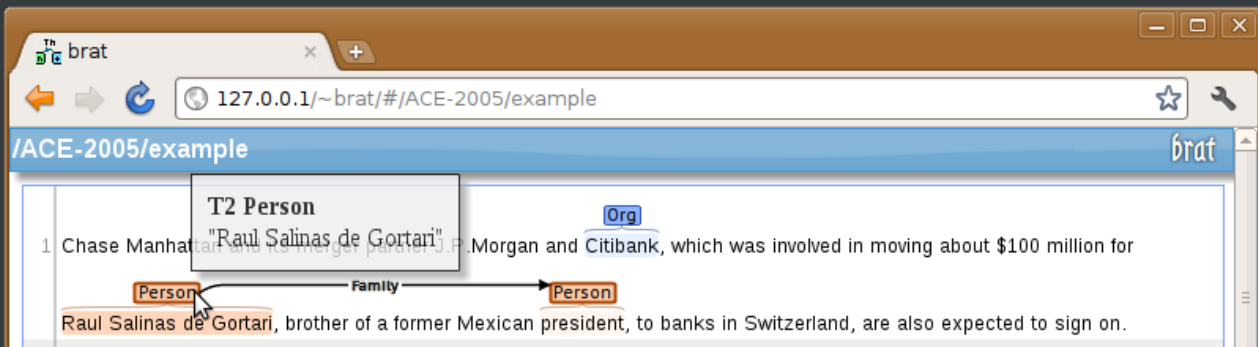
关系类型对话框

此对话框中可供选择的类型取决于集合的配置和两个选定实体的类型。

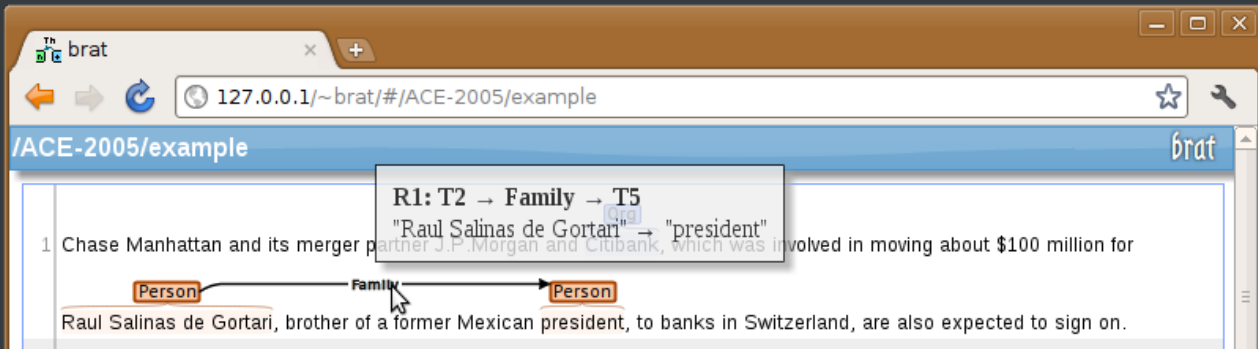
(当尝试关联没有任何可以连接它们的已配置关系类型时，将显示错误消息而不是此对话框。)

## 编辑现有的标注

登录用户可以通过双击标注来修改或删除现有标注。

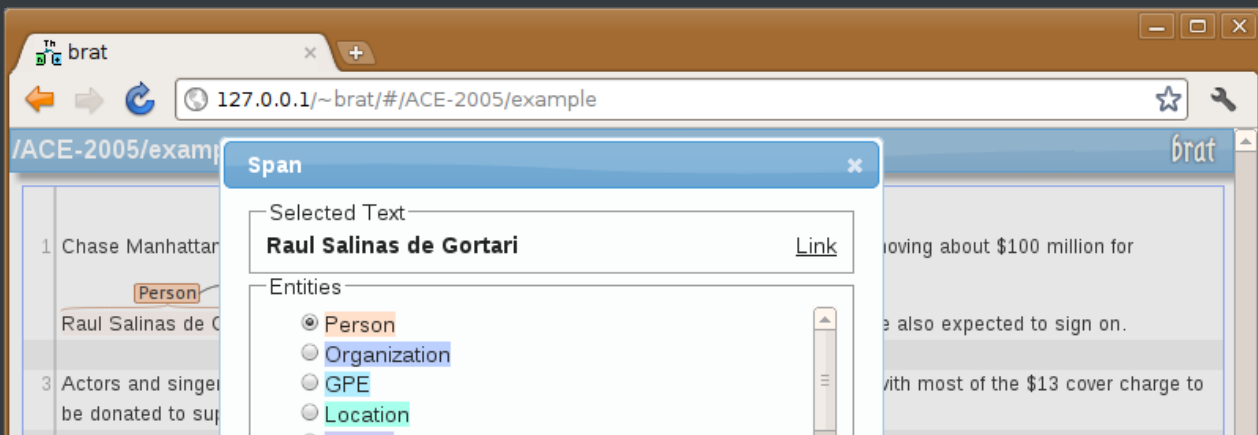


双击文本范围标注的"box"来进行编辑



双击连接两个标注的"arc"已编辑这些标注之间的关联

这将显示与最初选择标注类型相同的标注对话框，以及一些其他选项。



先前创建的实体标注的标注对话框

先前创建的关系标注的标注对话框

除了允许修改标注的类型和修改标注等其他方面外，这些对话框还支持以下附加功能：

- **Link:** 直接连到标注的URL
- **Delete:** 删除标注
- **Reselect:** 重新选择标注的范围或连接的标注

（注意，一旦创建，标注的基本类别就不能更改：也就是说，虽然实体标注可以更改为另一种类型的实体标注，但不能将其更改为事件或关系标注。）

## 归一化(Normalization)

有关如何使用归一化功能的信息（从brat v1.3开始可用），请参阅brat归一化页面（<http://brat.nlplab.org/normalization.html>）。

## 快捷键

在正常的可视化或标注状态下可以访问以下键盘快捷键。（当菜单打开时，这些键盘快捷键被禁用。）

- **ESC:** 清除所有打开的菜单（取消可能的修改）和屏幕消息
- **TAB:** 打开文件夹浏览器
- **CTRL-F:** 检索
- **CTRL-G:** 下一个搜索结果
- **Left arrow:** 上一个文档
- **Right arrow:** 下一个文档

按箭头键时访问“上一个”和“下一个”文档的顺序取决于在文件夹浏览器中选择的当前排序顺序。

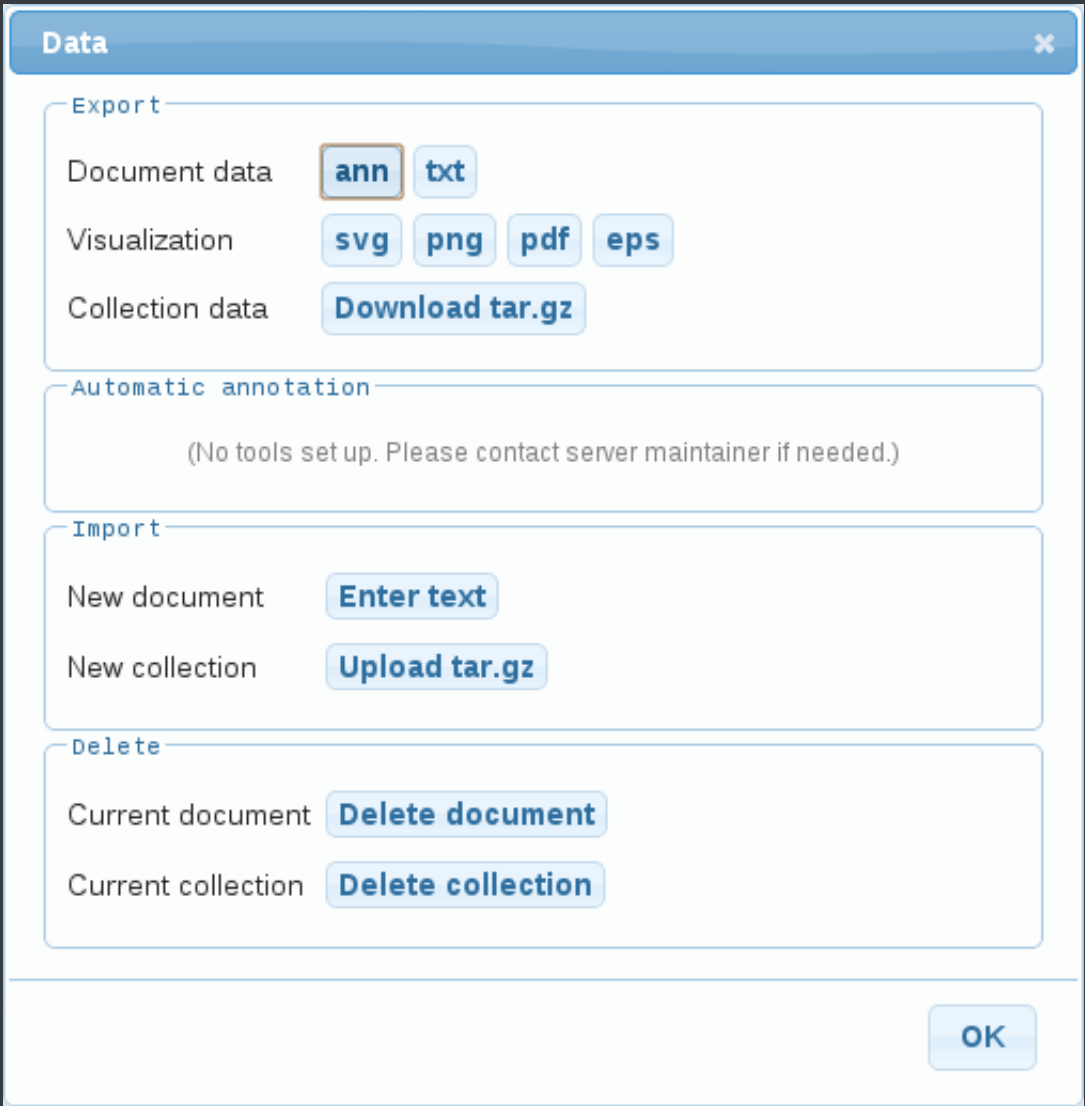


这允许使用键盘浏览文档，例如从具有最多实体标注的文档到具有最少实体标注的文档。（默认顺序是文档名称。）

打开标注菜单时，可以使用一组不同的可配置键盘快捷键来快速选择标注类型。

## 数据对话框

可从菜单访问数据对话框，提供对数据管理功能的访问。



### 数据对话框

数据对话框中包含以下功能：

- **导出文档数据：** 导出当前文档数据
  - **txt:** 原始文档文本 (UTF-8 encoding)
  - **ann:** brat stand-off 格式 的标注(text-based)
- **导出可视化文件：** 导出当前文档的可视化文件。支持的格式如下：
  - **svg:** Scalable Vector Graphics (vector format, XML)推荐进一步编辑的可视化文件(可以使用工具inkscape)

- **png:** Portable Network Graphics (bitmap format, binary) 适合网页图形
- **pdf:** Portable Document Format (vector format, binary) 适合嵌入文档
- **eps:** Encapsulated PostScript (vector format, binary) 适合嵌入文档
- **导出集合数据:** 导出一个用 gzip 打包的 tar 存档，其中包含整个集合的原始文本和注释。
- **自动标注:** 使用为 brat 安装配置的工具自动标注当前文档。
- **导入:**
  - **新文档:** 通过在简单对话框中输入文本，将单个文档添加到当前集合
  - **新集合(文件夹):** 使用gzip压缩的tar文件上传建立新的集合，其中可以包含生语料，也可以包含标注ann文件
- **删除:**
  - **当前文档:** 永久删除当前文档及其标注集合
  - **当前集合:** 永久删除当前集合，包括其所有文档及其标注

请注意，大多数这些功能仅在登录时可用，有些功能仅对具有管理权限的用户可用。

如果数据对话框中无法访问某些必要功能，请确保您已登录。如果问题仍然存在，请联系您的brat安装管理员。

## 搜索

### 基础搜索选项

可以从菜单或按CTRL-F访问的数据对话框提供的搜索功能。

#### 搜索对话框

顶部的四个选项卡允许选择要搜索的内容：

- **文本:** 搜索与给定字符串或正则表达式匹配的文本
- **实体:** 搜索与给定规范匹配的实体标注
- **事件:** 搜索与给定规范匹配的事件标注
- **关系:** 搜索与给定规范匹配的关系标注

默认情况下，唯一显示的选项是选择搜索范围：

可配置的搜索选项：搜索的范围是当前文档还是当前集合

- **Document:** 仅搜索当前文档
- **Collection:** 搜索当前集合中的所有文档

### 高级的搜索选项

单击“显示高级”文本可显示高级搜索选项。

## 高级搜索选项

- **语料库检索：**在上下文格式的关键字中显示具有指定大小（上下文长度）的周围上下文的搜索结果
- **匹配：**在匹配的给定文本字符串之间选择整个单词，单词的子串和正则表达式（正则表达式）搜索。在正则表达式模式下，给定字符串被解释为与目标文本匹配的（PCRE）正则表达式。
- **案例：**是否要求字符大小写在给定字符串和匹配文本之间保持一致。