

使用手册中文译文 (Chinese translated version for the readme):

AntConc3.2.1 使用手册 (Windows, Macintosh OSX, and Linux 操作系统)

#####

Laurence Anthony, Ph.D.(哲学博士)

科学工程英语教育中心

科学工程学院

早稻田大学

3-4-1 大久保, 新宿区, 东京 169-8555, Japan

3 月 10 日, 2007 年

#####

AntConc 开始是作为相对简单的索引程序, 不过慢慢地进步成一个相对有用的文本分析工具。它是在 ActiveState 供应商的优良 Komodo 跨平台支持多种程序软件的开发环境下用 Perl5.8 程序语言写成的。这个程序只要双击可执行文件就可以打开, 这个文件可以从 Laurence Anthony 实验室网站下载。这个程序可以在任何窗口环境下运行, 包括 Win 98/Me/2000/NT and XP, and also Macintosh OSX and Linux 的电脑。如果用户发现任何这个程序在一个特定的操作系统下的使用问题, 请让我知道。

AntConc 包括以下工具:

索引**Concordance**

索引定位**Concordance Plot**

文件查看**File View**

词丛**Clusters**

N 元模式 (部分词丛) **N-Grams (part of Clusters)**

搭配**Collocates**

词单**Word List**

关键词单**Keyword List**

注意每个工具都可以点击工具窗口上各自的选项卡或者使用功能键 F1 至 F7 打开。

“索引” (Concordance)

索引工具生成用户选择的一个或多个目标文件的索引行 (concordance lines 或上下文关键词: key word in context) 行。

concordance 使用步骤

1) 从 file 菜单的 open file (打开文件) 或 open dir (打开目录) 选择一个或多个要处理的文件, 选出来的文件按顺序在主窗户的左边框里显示出来。

2) 在左边的按钮条的输入框里输入一个检索词

3) 使用右边"Search Window Size" (检索窗口大小) 的按钮条的增加和减少按钮来选择在检索词两边显示的字符数。

4) 按 “Start” (开始) 键开始产生索引行的检索结果。检索过程中可按 “stop”(停止) 键随时停止检索。

5) 使用 Kwic Sort (上下文关键词分类) 下的按钮条选择一个目标词来重排索引行, 0 是检索词, 1L, 2L 是检索词左边的第一, 第二个单词, 1R, 2R 是检索词右边第一, 第二个单词。注意, 三级分类均可, 软件刚启动时, 二三级未选择。

6) 按 “Sort” (分类) 键开始分类处理。

7) 将指针移到其中一行索引行的突出的检索词之上, 系统默认为蓝色。指针会转变成一个手形的图标。点击突出的检索词, 可以使用户看到检索词在原文中出现的情况。见 “File View” (文件查看) 工具。

注意: 索引行的总数在按钮栏中间的 Concordance Hits (索引条数) 下的输入框里显示出来, 处理结束时, 这个数字会闪一下 “FINISHED” (结束), 如果某个检索词没有产生索引行, 则显示 “NO HITS” (零条), 如果是后面这种情况, 索引行的窗户就不会更新, 前面显示的索引行还会保留。

检索词可以通过选择检索词输入框（Search Term）上面的 word（单词）选项设定为“词（默认）”或“词的片断”，也可以通过选择 case 来决定区不区分大小写，不区分为默认，也可以选择“Regex”使用完整的正则表达式，可以参考很多有关正则表达式的文章。可以通过网上的搜索引擎来查。也可以在网站 <http://www.regular-expressions.info/quickstart.html> 上找信息。

按一下检索词输入框右边的 Advanced 键，可以进行更为复杂的搜索。第一个高级搜索项是定义一组检索词，既可以一行一个地输入，也可以直接载入文件中的一个检索词单。注意每一行都被认为是一个单独的检索词。这个特征允许用户使用一大组检索词但不用每次重复输入。第二个高级搜索选项是定义上下文词（Context Words）和一个上下文的范围，在这个范围内必须出现检索词。比如：要搜索"student"这个词，要求至少在它前三个或后三个单词中一定要出现"university"，就把检索词定为"student"，上下文词定为"university"，上下文范围定为‘From’3L‘to’3R

"索引定位"(Concordance Plot)

进行索引定位的步骤与索引工具的使用方法一样。但是索引定位工具提供了索引行另一个不同的视角。每一个文件的所有索引条数都在一个以条形码上定位出来，表示文中有检索词的索引行的位置所在。这个工具提供了一个容易的方法来看哪些文件包含目标检索词，也可以用来确定检索词在哪儿碰到词丛。一个例子是确定某个特定的实义词在科技文献的哪里出现，又或者一个字符什么时候出现在小说或剧本的发展过程中。

索引条数和每个文件的长度在条形码右边显示出来，定位本身的尺寸可以用左下角的 Plot Zoom 按钮来增大或缩小。

“文件查看”(File View)

任何时候目标文件都可以使用“文件查看”工具来查看原文。

要查看原文，用户需要进行以下操作：

- 1) 在主窗口的左边框里选择一个要查看的文件
- 2) 如果已经指定了一个检索词，检索词将会突出显示（默认蓝色）。选定主窗口上的 File View 按钮，再按开始（Start），则得到检索词在原文出现的情况。
- 3) 使用左下角的 Hit Location 下的按钮，可以跳到上一个或下一个检索词上。
- 4) 改变检索词点击开始键查看其它的检索项。
- 5) 点击在原文中突出显示的检索词，主窗口会出现一组 KWIC lines(上下文关键词行或索引行)

以下是一个“File View”的快捷键

Ctrl-Click（单击）=跳到窗口中最近的检索项上

“词丛”(Clusters)

词丛工具是用来生成目标文件检索词周围排列成序的词丛列单的。

词丛既可以按照频率也可以按照词头或词尾排序，也可以按照词丛第一个单词的出现概率排序。所有的排列顺序可以反向。用户也可以选择每个词丛的最大或最小长度（单词数），以及列出词丛的最小频率。也可以选择检索词是否总出现在词丛左边或右边。（注意：在目前的版本中，如果超过一个单词指定为检索词，选择检索词在右，只有第一个单词出现在右边。）

生成词丛列单，需进行以下操作：

- 1) 选定主窗口上的 Clusters 按钮，选择合适的排序选项。
- 2) 按开始键，任何时候可以使用停止键（Stop）中止。

3) 点击词丛会产生一组上下文关键词行或索引行。

“N 元模式”(部分词丛)

“N 元模式”工具是用来生成目标文件有序的 N 元列单的。N 元模式是单词的 N 元，因此，大的文件将会产生大量的 N 元，比如，一个句子 “This is a pen.” 中的二元是 ‘this is’, ‘is a’ 和 ‘a pen’。

就像和“词丛”工具一样，N 元模式工具既可以按词频也可以按词头或词尾排列。也可以按词丛首单词的出现概率排列。所有的列单顺序都可以反向。同样，用户也可以选择每个 N 元的最大或最小尺寸(单词数)，显示 N 元的最小词频。

想要列一个“N 元模式”的单子，用户需进行以下操作：

- 1) 选择 Clusters 窗口模式下 Search Term 右边的 N-Grams 选项
- 2) 选择合适的排列选项
- 3) 按开始键。任何时候都可以中止。
- 4) 点击词汇束会产生一组上下文关键词行。

“搭配”(Collocates)

“搭配”工具是用来生成目标文件的一组有序的搭配列单的。

搭配既可以按词频，也可以按照检索词左边或右边的搭配频率，或者词头或词尾排列。也可以按照检索词和搭配间的统计量的值来排列。统计量的值测的是检索词与搭配度之间的相关程度。AntConc 目前可行的统计量列在下面。所有的列单顺序可以反向。同样的，用户也可以选择检索词左右两边的单词跨度(单词数)，显示的搭配的最小频率。比如：如果只要求一个单词跨度，要看哪些单词直接出现在检索词右边，选中右边 Window Span(窗口跨度) 的 Same(相同) 选项，使得单词跨度的最大值和最小值相同。

统计量(Statistical Measures)

交互信息(MI)：使用 M. Stubbs 在 Collocations and Semantic Profiles, Functions of Language 2, 1 (1995) 此文中描述的方程式。

T 值(T-Score)：使用 M. Stubbs 在 Collocations and Semantic Profiles, Functions of Language 2, 1 (1995) 此文中描述的方程式。

要列一个搭配的列单，用户需要进行以下操作：

- 1) 选择合适的排列选项
- 2) 按开始键。任何时候都可以中止。
- 3) 点击搭配会产生一组上下文关键词行。

“词单”(Word List)

词单是用来生成目标文件的有序的词单(单词表)的。

词单既可以按词频，也可以按词头词尾的顺序排列，还可以反向。词单可以在大小写不敏感(case-insensitive)模式下生成，即不区分大小写，大小写同等对待；也可以在大小写敏感 case-sensitive 模式下生成，即区分大小写，大小写分别处理。

要列一个单词表，用户需要进行以下操作：

- 1) 选择合适的排列选项
- 2) 按开始键。任何时候都可以中止。
- 3) 点击单词会产生一组上下文关键词行。

“关键词单” (Keyword List)

除了生成普通词单, AntConc 可以比较目标文件中的单词和“参照语料库” (Reference Corpus) 中的单词来生成一个“关键词单”, 关键词单列出的往往是目标文件中频率很高或频率很低的词。

要列一个关键词单, 用户需要进行以下操作:

- 1) 选择一组目标文件
- 2) 到“Preferences”(预设)菜单中, 选择“Keyword Preferences”(关键词预设)选项。
- 3) 选择一个统计量估计目标文件单词的“关键性”(keyness), 推荐默认设定的对数相似值(Loglikelihood)。
- 4) 选择显示的关键词数的极限值(threshold Value)。
- 5) 选择是否显示“负关键词”(show negative keywords), 即与参照语料库相比目标语料库中不同寻常的低频词
- 6) 选择一个文本文件的参照语料库 (.txt), 步骤与选择目标文件相同。
- 7) 参照语料库的地址目录将显示出来(如果适当的话), 参照语料库的文件列单将在参照语料库选项(Reference Corpus Options)下的窗口中显示出来。
- 8) 点击应用“Apply”(应用), 返回主窗口。
- 9) 选择生成关键词单的排列选项, 方式同生成词单步骤 1)
- 10) 按开始键。任何时候都可以中止。
- 11) 点击关键词会产生一组上下文关键词行。

“菜单选项” (Menu Options)

菜单选项分为三组, “文件”(File), “总体设置”(Global Settings)以及“工具预设”(Tool Preferences)。每组可用选项如下所述:

“文件”

这里的选项与将文件读取到 AntConc 以及将包含各种不同类型的文件写入硬盘有关。也有选项是把所有当前的设置输出到文件(Export Settings to File), 或者引入文件上的用户设置(Import Settings from File)。如果用户设置文件由于某些原因损坏了, 只要重启文件或使用“恢复默认设置”(Restore Default Settings)选项将程序还原到原始状态即可。

“总体设置”

这里的范畴影响 AntConc 的多重工具。

“文件设置” <File Settings>

文件设置范畴中, 用户可以选择显示完整文件路径>Show full pathname)或只是文件名。有个问题是 File Settings 下面还有一个选择文件格式的项, Default file type to use with "OpenDir"然后下面有显示了.txt, 右边有个小按钮, 指的是你从 File 菜单选项里选择 OpenDir(打开目录)这个选项, 显示的是你可以导入的文件格式是 txt 形式, 意思是可以从 OpenDir 选项中选择的是这个小按钮按下去之后选择的文件格式。OpenDir 选择的只有文件夹, 所以必须建立一个文件夹之后再选。

“标注设置” <Tag Settings>

“标注设置”范畴中, 用户可以选择显示或隐藏语料库文件中的任何标注。如果要隐藏标注, 标注的开始标记 Start Tag 和结束标记 End Tag 必须确定, 默认的是<>。

“通配符设置” <Wildcard Settings>

通配符设置范畴中, 用户可以编辑默认的通配符字符这样就不会跟检索词条不符了。比如:“或”(Or)通配符的默认字符是(一个竖杠的字符| }, 可以换成反斜线符号/。

“定义类符(词)” <Token (Word) Definition>

定义类符(词)范畴中, 用户可以选择哪些字符, 数字等等定义为“词”。比如, 有些情况下, 只有字母组合(Letters)被认为是词, 但是其它时候, 也许会有需要将数字(numbers), 破折号等等包含在词的定义中。AntConc 完全适用统一码(Unicode), 也就是说它可以处理任何语言数据, 包括所有的欧洲语言和

亚洲语言。因此，默认选项以广义设置字母组合。比如：如果使用日语编码，字母组合则包括所有的日语字符。用户也可以做他们自己的“词”定义。

有关统一码标准（Unicode standards），更多信息见：

<http://www.cs.tut.fi/~jkorpela/unicode/guide.html>

<http://www.unicode.org/>

<http://www.unicode.org/Public/5.0.0/ucd/UCD.html>

<http://www.unicode.org/Public/UNIDATA/PropList.txt>

<http://www.unicode.org/charts/>

“颜色设置” <Color Settings>

颜色设置范畴中，用户能够编辑用于显示结果和其它信息的颜色。

“字体设置”

字体设置范畴中，用户能够编辑用于显示结果和其它信息的字体类型，大小，以及样式。

语言编码

<Language Encodings>

AntConc 完全适用统一码（Unicode），也就是说它可以处理任何语言数据，包括所有的欧洲语言和亚洲语言。

AntConc 读取数据的语言（编码）在这里指定。比如：如果你工作的数据保存的是欧洲语言，通常编码为 iso-8859-1（默认）。另一方面，日语文本则通常编码为 Shiftjis。通过指定恰当的编码，所有语言数据都能在 AntConc 中得到正确处理。

“工具预设”

每个工具（除索引定位和文件察看）都有一个预设类别，可以对设置做精细修改。所有工具预设类别允许客户显示或隐藏不同的结果显示框。比如：

用户能够选择在索引工具显示窗口中隐藏文件名的显示框，所有的工具也可以选择将所有数据视为小写（Treat all data as lowercase）并按大小写分类（use case when sorting）。如果结果显示区分大小写，包括大写字母在内的单词在列单中将会居前显示。

“索引预设” <Concordance Preferences>

除了以上这些，可以进行以下设置：

可以不根据检索词左右的单词排列结果，而可以按照检索词左右的字母安排结果。这样就可以搜索拼写区别了。

检索词也可以选择在关键词行（或索引行 KWIC lines)隐藏，使得教师可以测验学生选择可能的词来填空。

注意：任何时候，显示结果的窗口正在出现结果时，按 Ctrl-X 键， 检索词会隐藏或出现。

“词丛预设” <Clusters Preferences>

这个工具除了以上描述的没有另外的设置。

“搭配预设” <Collocates Preferences>

除了以上的设置，可以在这里选择统计量。当前，可以选择的有两种统计量：交互信息(MI) 和 T 值(T-Score)，见上文。

“词单预设” <Word List Preferences>

除了以上所述，可以做以下设置：

可以从文件中加载一个“词目单”(Lemma List),这个词目单可以用来生成词目单，而不是词单。使用词目

单功能时，主窗口中的“词目单词形式”(lemma word form(s))一栏将会显示语料库中与每个词目有关的单词。

通过指定“词目标记”(lemma entry)，也就是后接'->'，再接上词目所对应的一个或多个单词用非类符(逗号，空格，以及其它符号等)隔开，可以做一个词目单。例子如下：

be->is, are

play->play, plays, playing, played

注意以上的例子中，逗号和空格不认为是类符。因此，如果AntConc页面上使用了词目单(lemma list)，定义类符(词)(Token (Word) Definition)s设置中就需要加上破折号(dash)，以使词目单得到正确处理，因为带有连词符的合成词需要被看成一个词。

可以使用所有的词，或特定的一组词，

或忽略某一组词(停止词单)来生成词单。这被称之为“词单范围”(Wordlist Range)。将使用(或忽略)的单词范围可以直接由用户输入，也可以存储在文件里，然后按打开键(Open)由AntConc读取。用户也可以直接输入文件的单词组合和单词。

“关键词单预设”<Keyword List Preferences>

除了以上这些，可以进行以下设置：

如“关键词单”工具(Keyword List)中所述，要生成关键词单，用户需要指定一个参考语料库，以及一个关键性的统计量。尽管推荐默认的关键性统计量和极限值，可以在这个菜单里更改参数。选择“显示负关键词”(Show Negative Keywords)选项，与参考语料库相比目标语料库中不同寻常的低频词将显示出来。你也可以在这里交换主要语料库(目标语料库)和参考语料库(Swap Ref/Main Files)。

“快捷键”

这儿是一个快捷键的列单，可以用在所有使用窗口显示结果的工具上。

CTRL-C=复制当前选择的文本

CTRL-A=选择窗口的所有文本

ALT-A=选择所有窗口显示的文本(必须将指针移到窗口上)

双击=选择当前的单词

三击=选择窗口当前的一行

SHIFT-单击=选择所有窗口显示的连续行

CTRL-单击=选择所有窗口显示的不连续行

DELETE=删除所有窗口中选中的行

INSERT=除了选中的行，删除其它行

对于任何“spinbox”键(即标有上下可选择的一组键，如：检索词输入框边上的)，键盘上的方向键“上”(UP)和“下”(DOWN)可以用来启动上下的选择键。

“保存结果”

结果既可以保存在剪切板里，也可以保存到文本文件(..txt)或使用键盘命令，文件菜单上适当的选项或在每个

工具界面上分别点击“保存窗口”键(Save Window)保存到新的窗口。并且，也可以通过双击附加名为.exe的文件打开很多个相同的AntConc。

其它评论

所有的新的版本和漏洞修补都在下面的修改历史中列出来了。然而，如果用户发现这个程序有漏洞，或者对于改进程序有任何建议，请让我知道，我会在以后的版本里改进这些问题。确实，大部分是由于世界各地的用户的评论，做出的修改。对此我深怀感激。

这个软件可以用来作为“免费软件”使用（见法律事务），但是很重要的是我的经费提供者希望得知任何人使用这个软件获得成功的情况。因此，如果你们发现这个软件很有用，请给我发电子邮件简单地告诉我它是如何使用的。

致谢

我想要对 AntConc 的使用者说谢谢，他们不厌其烦地给我发电子邮件告诉我关于软件的反馈信息，并且提出提高和改进的建议。AntConc 的开发由日本教育科学运动文化提高社团的科研财政补贴(No. 16700573)以及日本早稻田大学专项研究项目财政(No. 2004B-861)提供支持。

法律事务

AntConc3.1.302 可以自由用于个人非营利性使用，如果附加这个使用手册不做任何改变的话，可以自由传播。如果这个软件计划用在团队环境中，需要通知我这个软件将如何使用，然后我将决定你是否得到使用许可。这个软件以现状运行，如果用户使用软件造成损害，本软件无赔偿义务。