

# 硕士学位论文

## 基于汉民族思维模式的情感计算

Sentiment Analysis in Chinese Thinking Model

作者姓名：\_\_\_\_\_王宇轩\_\_\_\_\_

学科、专业：\_\_\_\_\_计算机应用技术\_\_\_\_\_

学号：\_\_\_\_\_20809318\_\_\_\_\_

指导教师：\_\_\_\_\_林鸿飞 教授\_\_\_\_\_

完成日期：\_\_\_\_\_2010. 11\_\_\_\_\_

大连理工大学

Dalian University of Technology

## 大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目： \_\_\_\_\_

作者签名： \_\_\_\_\_ 日期： \_\_\_\_\_年\_\_\_\_月\_\_\_\_日

## 摘 要

情感倾向性分析是情感计算中的一个重要领域。本文从对词与句的情感分析开始，通过对汉民族和西方民族思维方式差异的研究，针对基于词典的情感分析的方法给出了考虑思维模式后的篇章情感倾向性分析方法。

本文以 COAE2009（中文倾向性分析评测）中的评测任务一以及任务二为目标，对作为篇章的基本组成单位的词以及句子进行分析，利用大连理工大学情感词汇本体库作为主要的情感词汇来源，完成对情感词的识别，同时针对句子中的情感词，考虑了句子中的程度副词、否定词、转折词以及句子长度的影响因素给出了句子的情感倾向性的分析。从评测的实验结果上看，本文的方法能够对情感表达明显的句子进行优先识别，在评测指标 P@1000 上高出了评测平均水平 16 个百分点。

针对篇章级别的情感倾向性分析工作，本文利用了 COAE2009 中的相关工作成果，在对中文文本倾向性的研究中发现，中文文本在表达情感时，不仅受到语法结构以及修饰成分的影响，同时还受到中华文化的影响，而这种影响在目前的研究中尚未展开深入的研究。本文针对汉民族的“图形式”、“具体性”、“散点视”的思维特点进行研究，同时又通过与西方民族“直线式”“抽象性”以及“聚点视”思维特点的比较，给出了初步量化思维的方法。对于情感分析过程中不能明确表达情感倾向性的文本，本文使用 LSA 方法，利用情感表达明确的文本形成语义网络，检索与情感表达不明确文本的语义相似度高的情感表达明确的文本，来确定情感表达不明确文本的最终情感倾向。实验结果表明，在基于词典的篇章情感分析方法中考虑思维模式的影响因素后，情感倾向性的分析效果有了提高，进而证明了中文情感表达中，思维因素是影响文本倾向性的一个因素，并且证明了本文提出的量化方法的有效性。

**关键词：**汉民族思维模式；LSA；情感倾向性分析；情感计算



## Sentiment Analysis in Chinese Thinking Model

### Abstract

Sentiment Analysis is an important part of the Emotion Computing. Based on researches on sentiment analysis of words, sentences and the differences between Chinese thinking model and Western thinking model, we proposed a method, which was a sort of lexicon-based method, which could attain the emotion of articles.

To accomplish COAE2009's (Chinese Opinion Analysis Evaluation) task one and task two, we did some researches on the words and sentences which are the basic elements of an article from scratch. Meanwhile we took full advantage of Dalian University of Technology Affective Lexicon Ontology to distinguish the emotional words from others and discerned the emotion of a sentence by the emotional words, adverbs, negative words, sentence length and but-clause. The evaluational result of our achievement shows that our method gave the priority to strong-emotion sentences, and achieved an advance that is higher than average P@1000 by 16 percents.

During the researching of sentiment analysis on articles, we discovered that the factors which could influence the emotions of articles were not only the decorative elements but also the thinking model of people, which were not paid attention by other researchers. We proposed a set of methods to compute the Chinese thinking features such as graphical model, specific model and scattering model compared with Western thinking models such as straight model, abstract model and focusing model. As to the weak-emotion sentences, we use LSA methods to clarify the emotion of uncertain samples by retrieving the seed samples in the semantic space. The experimental results showed that the performance of sentiment analysis involved with thinking models was much better than the lexicon-based methods. So we could make a conclusion that thinking model was an important factor in researches on sentiment analysis and our method was an effective method to the sentiment analysis.

**Key Words:** Chinese Thinking Model; LSA; Sentiment Analysis; Affective Computing



## 目 录

摘 要.....	I
Abstract .....	III
1 绪论.....	1
1.1 研究背景.....	1
1.2 研究现状.....	2
1.3 论文的工作.....	4
1.4 本文的结构.....	5
2 情感词句的识别以及分类.....	6
2.1 相关知识以及资源介绍.....	6
2.1.1 任务介绍.....	6
2.1.2 情感词汇本体库.....	7
2.2 任务实现.....	10
2.2.1 情感词的识别及其分类.....	10
2.2.2 情感句的识别及其分类.....	10
2.3 结果分析.....	13
2.4 本章总结.....	15
3 汉民族思维模式下情感分析.....	16
3.1 相关知识.....	16
3.1.1 中西思维特点及其差异.....	16
3.1.2 LSA 介绍.....	18
3.2 汉民族思维模式计算方法.....	20
3.2.1 基于思维共性的情感计算.....	20
3.2.2 中文情感表达模式.....	21
3.2.3 基于 LSA 的隐式表达挖掘.....	23
3.3 实验语料.....	25
3.3.1 实验设计.....	25
3.3.2 实验分析.....	27
3.4 本章总结.....	27
4 中西思维差异对情感计算的影响.....	29
4.1 中西思维差异及量化方法.....	29
4.1.1 图形式与直线式.....	29

4.1.2 具体性与抽象性.....	29
4.1.3 散点视与聚点视.....	30
4.2 实验结果与分析.....	31
4.2.1 实验语料集.....	31
4.2.2 实验设计.....	31
4.3 本章总结.....	35
结    论.....	36
参    考    文    献.....	37
攻读硕士学位期间发表学术论文情况.....	40
致    谢.....	42
大连理工大学学位论文版权使用授权书.....	43



# 1 绪论

## 1.1 研究背景

现实世界中的文本信息类型分为事实类型和观点类型，事实类型的文本是对事件或其属性的客观描述，而观点类型则是作者对事件或其属性的主观观点或感觉。随着 Web2.0 时代的全面来临，用户产生的评论以产品评论，博客以及讨论组等方式大规模地在网络上流行，这些主观情感信息提供了给我们诸多的应用，例如：产品评价系统等。在针对观点类型的文本的研究中，情感计算成为了其中一个重要的研究方向。

情感计算属于人工智能方面的一个热门研究领域，其领域的研究的核心目标是使计算机拥有像人类一样的情感，美国 MIT 实验室的 R.Picard 教授早在 1997 年在其的专著“*Affective Computing*（情感计算）”一书中将其定义为“情感计算是与情感相关，来源于情感或者能对情感施加影响的计算”<sup>[1]</sup>。情感计算包括两方面的研究，即首先使得计算机能够识别人类的情感，其次是使得计算拥有像人一样的表达自身的情感。

文本作为网络主要的信息载体得到了研究人员的重视，同时文本倾向性分析作为情感计算的一个重要分支同样得到了研究人员的关注。现如今网络的普及促使电子商务突飞猛进的发展，大量的产品信息以及产品评价信息发生在互联网上，其中产品的评价信息成为生产商以及用户所使用的重要信息。生产商针对用户的使用感受以及对产品的优缺点的评价，可以针对缺点做出相应的改进，而用户根据网页上的评价行信息来决定是否购买产品，不会像以前那样通过询问好友的用户体验来做决定，即可以节省时间，又得到了更多的信息。所以针对评论内容的情感倾向性分析有着重要的研究意义。

网络上流行的另一种文本形式是博客。人们除了对亲朋好友进行情感的倾诉外，更倾向于将自己所经历的事情以及所产生的情感以博文的方式发表到网络上，读者针对博文中描述的事件以及抒发的情感给出相应的回复，博文的作者则能够通过网友的劝解获得心灵的解脱，或是找到相同经历的人的回复，产生共鸣。所以对博文的情感分析成为了一个新的应用。随着 Internet 的发展，观点类型的文本的增多，为我们研究情感计算提供了实验的数据，从海量文本中提取其中包含的情感信息，在许多方面都有广阔的应用，如邮件系统、个性化文本、网页评价等。现在主要的研究成果有 EmpathyBuddy 邮件系统<sup>[2]</sup>，彩条化显示文章的情感<sup>[3]</sup>，名人的评价<sup>[4]</sup>等。

国外学者较早地针对情感计算进行了研究，主要是针对英文语料提出了提高情感计算效果的算法，加之国外网络较中国提前发展起来，网络上存在大量的英文文本的信息，为其提供了良好的数据来源，然而国内尚缺乏相关的资源，故尚未得到成熟的算法。随着网络时代信息量的急剧增加，中文的用户评价以及情感表达文本的数量也随之增加，

国内关于情感计算的研究也逐步发展壮大,趋于成熟,但相比较国外的研究还有所差距,在这种背景下,本文的研究中心为以中文语料为研究对象,从思维决定语言表达方式这个认识出发,以提高中文情感倾向性分析效果为目标,提出了改进情感倾向性效果的方法。

## 1.2 研究现状

情感倾向性计算作为情感计算的一部分,其主要任务为识别出文本所要表达的情感倾向,即得到作者对事物的态度以及作者通过文本要表达的喜怒哀惧的情感。针对文本情感倾向性研究,目前国内外主要研究包括以下两种:一种为定位关键情感词方法,即在文本中寻找情感词的位置,通过情感词的情感类别来确定文本的情感类别,其他类似的方法是平均化句子中情感词汇的方法,即根据句子中的情感词的平均值来确定句子情感<sup>[5]</sup>;另一类为基于统计的自然语言处理的方法,该类方法主要是采用机器学习方法,例如神经网络<sup>[6]</sup>,基于 HMM 的情感建模方法<sup>[7]</sup>等。通过对上述方法的调研发现,这些文本情感计算方法都依赖于情感词汇,如果情感句中不包含情感词汇,上述方法的效果就不理想了,由此看出情感词汇成为文本倾向性研究中主要使用的资源。在众多词典中,大连理工大学情感词汇本体库<sup>[8]</sup>根据目前情感分类方法研究的现状,明确了情感分类体系,并由此基础上充分利用现有的情感词汇资源构建情感词汇本体,采用手工分类以及自动获取相结合的方法扩充词汇本体,记录了词汇的情感类别、强度和极性等信息。

除了以情感词为主体的方法外,还可以利用其他资源进行情感倾向性分析。Hugo Liu<sup>[2]</sup>等采用常识知识库资源对文本进行情感倾向性分析,文中对于那些不包含情感词汇的情感句的情感倾向性是通过大规模的常识知识库获得的。当采用的常识知识库的规模以及覆盖的知识领域范围足够广,可以只使用常识知识库的资源对文本进行情感计算,但是该类方法需要有较为完善地知识库体系来保证分析的准确性。

除了上述的方法外,还有基于非监督学习的方法,文献[9]在评论性文本中计算评论中的短语与"excellent"和"poor"的互信息值,将该值作为短语的观点倾向值,如果一条评论中短语的互信息平均值倾向于积极的,那么该条评论则别认为是积极的,反之则被认为是消极的评论。文献[9]在汽车,银行,电影以及旅游相关的410条评论上达到了平均准确率为74%。

按照文本级别的分类分为词汇级别,句子级别以及篇章级别的研究。文献[10][11][12][13][14][15][16]都是针对词汇级以及句子级情感识别的研究。文献[10]认为词汇级的特征是识别分析句子级情感的关键,该文以层级为前提假设用以划分词汇特征及其相关关系,使用分析工具进行分析以减少特征集合数量达到最优特征集的作用,并在

三种分类任务上均达到较好的效果。而在文献[11]中根据语言学方面的要求，使用 **bootstrapping** 方法充分抽取表达出句子情感的模板，这种方法可以抽取出有效的情感模板并得到较高的召回率。同样，统计学方法如朴素贝叶斯结合相应的平滑方法等也可以适用于句子级的情感识别以及分析中，亦在分类任务中取得了很好的效果<sup>[12]</sup>。文献[15]中的方法与传统的处理方法有所不同，不但考虑了传统的特征集合，并将句子长度和代表情感的语法级特征相融合，并在实验中得到了很好的验证。

句子级别的情感分析方法中基于词典的情感计算方法，都需要构建相应的情感词典，其中通过人工方法构建的有[17][18][19][20]。但是这种方法不是被单独使用的，其中基于情感词典分析扩充词典的研究有[18][20]。在文献[18]，使用语法规则和语言学规则初步过滤情感句，然后通过抽取特性词、共现词、典型句子特征和对相应类别的一致性分析特征这四类，判断句子的极性，最终得到产品的声誉度。而在文献[20]中，通过情感词典和情感模板数据库，可以对在线的文档抽取主题特征集，情感特征集以及情感主体间关系抽取，目前在产品评论方面有很好的应用。文献[21]采用的是种子词，利用 **wordnet** 等增加词典内容，该种方法的缺点在于不能有效得解决领域倾向性词汇问题，例如电话的“无声”为 **negative**，汽车的“无声”为 **positive**。

目前国内的研究大部分的研究工作集中在词汇级别以及句子级别上。主要表现在国内的评测任务设置上，将任务集中安排在了情感词以及情感句的识别上，例如 **COAE2008**<sup>[22]</sup>（第一届中文倾向性分析评测）中的任务一要求自动识别出测试集中包含的情感词，即在一定的上下文环境中能够明确表达观点倾向性的词语。任务二要求对任务一识别出来的情感词进行褒贬极性判别，即褒义、贬义。输出任务一识别出的所有情感词的褒贬极性判别结果并按照置信度排序。**COAE2009**<sup>[23]</sup>中的任务一要求自动识别出测试集中包含的情感词并分类，即在一定的上下文环境中抽取出能够明确表达人物情感的词语，并判断该情感词所属的类别，以及任务二要求自动识别出测试集中包含的情感句并分类，即在一定的上下文环境中抽取出能够明确表达人物情感的句子，并判断该情感句所属的类别。针对情感词的研究，谭松波<sup>[24]</sup>针对不同领域的特点，建立起来领域内文本的词与词之间的语义网络，词与文档之间的语义网络，以及词与领域外的文档的语义网络，最终得到适用于目标领域情感词典。针对不同领域间的情感文本分类问题，文献[25]根据源域和目标领域的文本相似关系，建立了两个领域的语义关系网络，最终得到新领域内的文本的情感倾向性。目前篇章级的倾向性研究较少，原因在于篇章级文本的情感表达、结构以及评论主体的多样性。篇章级情感倾向性分析，是从整体上判断某个文本的情感倾向性，即褒贬态度以及作者的情绪。有代表性的工作包括Turney对电影评论的分类<sup>[26]</sup>。Turney的方法是将文档中词汇和短语的倾向性进行平均，来判断文章的

倾向性。这种方法基于情感倾向性词典，不需要人工标注文本情感倾向性的训练语料。Liu Bing等人基于整体词典的方法进行观点挖掘<sup>[27]</sup>，是对传统方法的改进。基于词典的方法过于依赖词典，需要良好的词典做支持才能有很好的结果，谭松波等人针对不同领域建设词典<sup>[24]</sup>虽然能够提供较好的词典，但针对不同的领域要先建立相应的词典，才能保证一定的准确性，缺乏灵活性。徐琳宏的基于语义资源的文本情感分析<sup>[28]</sup>，是利用CRF对文章逐句进行情感标注，得到文章的情感链，进而判断文章的倾向性。目前关于篇章级别的情感倾向性工作不是很多，但是网络文本的大量出现，以及博文成为主要的情感表达的方式以后，篇章级别的情感分析工作越来越需要被重视。

### 1.3 论文的工作

本文根据COAE2009中的任务一以及任务二的相关问题，初步实现了对情感词以及情感句倾向性的分析工作，为针对篇章级别的文本情感倾向性的分析工作奠定了基础，同时完成了对基于词典的方法的改进。本文在对情感倾向性的研究中发现汉民族的情感表达深受中华文化的影响，而这种影响因素在调研的文献中并没有给予考虑，而东西方文明是两种不同的文明，不同的文明造就了东西方认知上的不同，认知上的差异又导致了东西方在情感表达上的不同，目前，在情感计算的研究方法中多数方法是依据英文的文本进行的研究工作，这些方法不完全符合东方民族尤其是汉民族的思维，不考虑思维差异而直接使用适用于英文语料的方法，可能会造成方法的失效。为了能够更好得完成对中文文本的情感分析，以及能够更好得利用英文的情感计算分析方法来实现对中文文本的情感分析方法的改进，本文考虑思维的差异性，情感表达的特殊性，将上述因素应用到中文情感倾向性分析中，改进中文情感倾向性分析的效果。

本文从研究中西文化开始，以基于词典的方法为基本的实验方法，充分考虑思维方式对该类方法的作用，从下属几个思维特点中，总结出提高基于词典的方法对情感分析的效果的方法。

#### “散点视”与“聚点视”思维方式

汉民族的思维的一个特点为“散点视”即在中文常使用多个相同含义的词汇来表达同一个含义，而西方民族的“聚点视”思维方式则着重使用一个能够清楚表达作者本意的词，所以在情感表达中多词的使用常使得中文显示出气势以及唯美的感觉，英文则更体现其重视逻辑，言简意赅的特点。

#### “具体性”与“抽象性”思维方式

“具体性”思维特点表示中文在表达中为了能够表达含义，常常采用形象的，具体的事物来表达，即利用两件事物之间的内在相似性进行表达，而在表达中不明确说出相

似的部分为哪些，具有借物寓意的意味。英文在表达时常使用抽象含义的词汇来表达抽象意义，不借助外物来表达，虽然在形式上缺乏了中文的形象，生动的特点，但是体现出了用语表达上的准确性，不会导致对待同一件事物，不同的人对“借物寓意”中的“意”有理解上的差别。

#### “图形式”与“直线式”思维方式

汉语具有委婉，含蓄的特点，这种特点在行文上表现为先描述与中心表达意思相关的事物，但不明确要表达的含义，往往在行文的末端处才表达出其真正的含义。英文“直线式”思维方式体现为表述过程往往开门见山地提出或者阐明自己的观点。

#### “人称式”与“物称式”思维方式

汉民族“人称式”的思维特点表现为中文常使用人作为句子的主语，而英文则常使用物即“it”作为主语。

总结出中西方四对思维特点后，本文通过对这些思维特点的分析，针对中文的“散点视”、“具体性”、“图形式”进行量化，加入到基于词典的方法中，“人称式”虽然是汉民族的思维的一个特点，但是在情感表达中所能发挥的作用并不明显，故本文并没有针对“人称式”的思维特点给出相应的处理方法。

## 1.4 本文的结构

本文共分为四个章节，主要内容如下：

第一章为绪论，主要介绍了研究的背景以及研究的现状，介绍了关于情感计算领域的国内外的相关研究以及研究现状。

第二章介绍了本文在 COAE2009 任务一以及任务二评测任务中对情感词以及情感句的识别工作，同时介绍了大连理工大学情感词汇本体库。

第三章介绍了汉民族思维模式下情感分析的方法，同时介绍了汉民族思维特点，将汉民族思维特点分为了显示特征和隐式特征，并针对各自的特征给出了相应量化的方法，方法的有效性在随后的实验中得到了验证。

第四章针对第三章的算法提出了细化研究，将思维特征中的“散点视”以及“具体性”思维方式进行了细化分析，同时给出了相应的量化思维方式的方法。并通过实验验证了方法的可行性和有效性。

最后针对课题工作进行总结及展望。

## 2 情感词句的识别以及分类

### 2.1 相关知识以及资源介绍

#### 2.1.1 任务介绍

本文根据 COAE2009 中的任务一以及任务二完成对情感词以及情感句的情感倾向性分析，下面给出 COAE2009 的相关任务介绍。

表 2.1 COAE2009 任务设置

Tab.2.1 Tasks in COAE2009

类型	任务说明	评测任务	任务号	组号
词汇级	中文情感词的识别及分类	情感词识别及分类	1	一
句子级	中文情感句的识别及分类	情感句识别及分类	2	
句子级	中文观点句子抽取	观点句抽取	3	二
要素级	中文观点倾向性相关要素抽取	观点评价对象抽取	4	
篇章级	面向对象的中文文本观点检索	观点检索	5	三

上表中显示了 COAE2009 所有的任务分配，本文详细说明任务一以及任务二的相关要求。任务一要求自动识别出测试集中包含的情感词并判断该情感词的所属的类别，主要分为四类情感：喜、怒、哀、惧，输出结果要求按照置信度降序排列，同时要求给出情感词所在的文章编号，以情感词为中心、前后各 20 个字节组成的文本片段，用于消除评测歧义。

表 2.2 任务一示例

Tab.2.2 Demonstration of task1

序号	单位标识	情感词	来源文本	情感词前后文本片段 (20 字节)	词置信度
1	abcd	高兴	Doc1	同学的肯定，我真的好高兴啊！！哈哈！不枉费我	1
2	abcd	开心	Doc1	，只要自信！我真的好开心，心都亮起来了哈哈！	1
3	abcd	伤心	Doc2	爱到深了，痛了，伤心，于是泪就有了	1
4	abcd	痛彻心扉	Doc2	的是一种无奈的选择，痛彻心扉。	1
5	abcd	泪	Doc2	痛了，伤心，于是泪就有了，哭是最好的	0.8
6	abcd	痛	Doc2	爱到深了，痛了，伤心，于是泪就	0.8
7	abcd	哭	Doc2	心，于是泪就有了，哭是最好的选择……完了	0.8
8	abcd	哈哈	Doc1	我真的好高兴啊！！哈哈！不枉费我那么辛苦的	0.8
9	abcd	无奈	Doc2	放手真的是一种无奈的选择，痛彻心	0.5

靡。

任务二要求自动识别出测试集中包含的情感句并判断该情感句的所属的类别，主要分为四类情感：喜、怒、哀、惧。输出结果要求按照置信度降序排列，同时要求给出情感句所在的文章编号。

表 2.3 任务二示例

Tab.2.3 Demonstration of task2

序号	单位标识	来源文本	情感词前后文本片段（20 字节）	句置信度
1	abcd	Doc1	今天的表现得到老师和同学的肯定,我 真的好高兴啊!!	1.0
2	abcd	Doc1	我真的好开心,心都亮起来了哈哈!	1.0
3	abcd	Doc1	哈哈!	0.8
4	abcd	Doc1	真是值得啊!	0.6

表 2.2 以及表 2.3 分别给出了任务最终提交的格式，为了完成对情感词的识别，本文需要一个基础的词典作为辅助，本文采用了[8]中的情感词汇本体库，在 2.1.2 章节中我们将详细介绍该情感词汇本体库。

### 2.1.2 情感词汇本体库

本文使用的情感词典为文献[8]构建的情感词汇本体库，其分类标准是在 Ekman<sup>[29]</sup>的六大类情感的基础上将情感划分为 7 大类，将“高兴”类更名为“乐”类，并为了增加对积极情感的细化分类，在原来的分类基础上增加了“好”一类来描述喜好、喜欢类型的情感，增加对积极情感类的刻画。[8]划分的情感 7 大类为乐，好，怒，哀，惧，恶，惊。为了便于在以后研究细致情感时方便使用，文献[8]在每个大类内按照情感强度和复杂度的区别程度，细化情感大类，分为 20 种小类情感，这 20 小类分别为：快乐、安心、尊敬、赞扬、相信、喜爱、愤怒、悲伤、失望、疚、思、慌、恐惧、羞、烦闷、憎恶、贬责、妒忌、怀疑、惊奇。下面举例“好”大类情感的分类来说明小类分类原则。“好”大类情感被分成了“尊敬”、“赞扬”、“喜爱”和“相信”四个小类情感，其分类的标准为这四类情感按照情感强度依次减弱的顺序进行分类。

表 2.4 给出了情感分类情况以及例词。

表 2.4 情感分类明细表  
Tab.2.4 Sentiment Classified

编号	情感大类	情感类	例词
1	乐	快乐	喜悦、欢喜、笑咪咪、欢天喜地
2		安心	踏实、宽心、定心丸、问心无愧
3	好	尊敬	恭敬、敬爱、毕恭毕敬、肃然起敬
4		赞扬	英俊、优秀、通情达理、实事求是
5		相信	信任、信赖、可靠、毋庸置疑、
6		喜爱	倾慕、宝贝、一见钟情、爱不释手
7	怒	愤怒	气愤、恼火、大发雷霆、七窍生烟
8	哀	悲伤	忧伤、悲苦、心如刀割，悲痛欲绝
9		失望	憾事、绝望、灰心丧气、心灰意冷
10		疚	内疚、忏悔、过意不去、问心有愧
11		思	相思、思念、牵肠挂肚、朝思暮想
12	惧	慌	慌张、心慌、不知所措、手忙脚乱
13		恐惧	胆怯、害怕、担惊受怕、胆颤心惊
14		羞	害羞、害臊、面红耳赤、无地自容
15	恶	烦闷	憋闷、烦躁、心烦意乱、自寻烦恼
16		憎恶	反感、可耻、恨之入骨、深恶痛绝
17		贬责	呆板、虚荣、杂乱无章、心狠手辣
18		妒忌	眼红、吃醋、醋坛子、嫉贤妒能
19		怀疑	多心、生疑、将信将疑、疑神疑鬼
20	惊	惊奇	奇怪、奇迹、大吃一惊、瞠目结舌

在情感词汇本体中，文献以下面的结构进行词汇的存储，即采用设计为一个三元组，如下所示

$$Lexicon = (B, R, E)$$

其中 **B** 表示词汇的基本信息，主要包括词汇编号、名称、英文译义、词性和版本信息等内容。**R** 以哈尔滨工业大学的同义词词林为主要参考，表示了词汇之间的同义关系，从同义词中人工挑选具有情感色彩的词汇录入，然后修改具有同义关系的一组词的