

中文情感词汇本体说明文档 V1.0

1. 介绍

中文情感词汇本体库是大连理工大学信息检索研究室在林鸿飞教授的指导下经过全体教研室成员的努力整理和标注的一个中文本体资源。该资源从不同角度描述一个中文词汇或者短语，包括词语词性种类、情感类别、情感强度及极性等信息。

中文情感词汇本体的情感分类体系是在国外比较有影响的 Ekman 的 6 大类情感分类体系的基础上构建的。在 Ekman 的基础上，词汇本体加入情感类别“好”对褒义情感进行了更细致的划分。最终词汇本体中的情感共分为 7 大类 21 小类。

构造该资源的宗旨是在情感计算领域，为中文文本情感分析和倾向性分析提供一个便捷可靠的辅助手段。中文情感词汇本体可以用于解决多类别情感分类的问题，同时也可以用于解决一般的倾向性分析的问题。

2. 本体格式介绍

情感词汇本体中，一般的格式为。

表 1 情感词汇本体格式举例

词语	词性种类	词义数	词义序号	情感分类	强度	极性	辅助情感分类	强度	极性
无所畏惧	idiom	1	1	PH	7	1			
手头紧	idiom	1	1	NE	7	0			
周到	adj	1	1	PH	5	1			
言过其实	idiom	1	1	NN	5	2			

其中，一个情感词可能对应多个情感，情感分类用于刻画情感词的主要情感分类，辅助情感为该情感词在具有主要情感分类的同时含有的其他情感分类。

3. 情感分类及情感强度

情感分类按照论文《情感词汇本体的构造》所述，情感分为 7 大类 21 小类。

情感强度分为 1,3,5,7,9 五档，9 表示强度最大，1 为强度最小。

情感分类如表 2 所示：

表 2 情感分类

编号	情感大类	情感类	例词
1	乐	快乐(PA)	喜悦、欢喜、笑咪咪、欢天喜地
2		安心(PE)	踏实、宽心、定心丸、问心无愧
3	好	尊敬(PD)	恭敬、敬爱、毕恭毕敬、肃然起敬
4		赞扬(PH)	英俊、优秀、通情达理、实事求是
5		相信(PG)	信任、信赖、可靠、毋庸置疑
6		喜爱(PB)	倾慕、宝贝、一见钟情、爱不释手
7		祝愿(PK)	渴望、保佑、福寿绵长、万寿无疆
8	怒	愤怒(NA)	气愤、恼火、大发雷霆、七窍生烟

9	哀	悲伤(NB)	忧伤、悲苦、心如刀割、悲痛欲绝
10		失望(NJ)	憾事、绝望、灰心丧气、心灰意冷
11		疚(NH)	内疚、忏悔、过意不去、问心有愧
12		思(PF)	思念、相思、牵肠挂肚、朝思暮想
13	惧	慌(NI)	慌张、心慌、不知所措、手忙脚乱
14		恐惧(NC)	胆怯、害怕、担惊受怕、胆颤心惊
15		羞(NG)	害羞、害臊、面红耳赤、无地自容
16	恶	烦闷(NE)	憋闷、烦躁、心烦意乱、自寻烦恼
17		憎恶(ND)	反感、可耻、恨之入骨、深恶痛绝
18		贬责(NN)	呆板、虚荣、杂乱无章、心狠手辣
19		妒忌(NK)	眼红、吃醋、醋坛子、嫉贤妒能
20		怀疑(NL)	多心、生疑、将信将疑、疑神疑鬼
21	惊	惊奇(PC)	奇怪、奇迹、大吃一惊、瞠目结舌

4. 词性种类

情感词汇本体中的词性种类一共分为 7 类，分别是名词（noun），动词（verb），形容词（adj），副词（adv），网络词语（nw），成语（idiom），介词短语（prep）。

5. 极性标注

每个词在每一类情感下都对应了一个极性。其中，0 代表中性，1 代表褒义，2 代表贬义，3 代表兼有褒贬两性。

注：褒贬标注时，通过词本身和情感共同确定，所以有些情感在一些词中可能极性 1，而其他的词中有可能极性为 0。

6. 存储格式及规模

中文情感本体以 excel 的格式进行存储，共含有情感词共计 27466 个，文件大小为 1.22M。