

《信息处理用现代汉语分词词表》规范*

孙茂松¹ 王洪君² 董秀芳²

清华大学智能技术与系统国家重点实验室¹ 北京大学中文系²

E-mail: sms@s1000e.cs.tsinghua.edu.cn

摘要: 本文给出我们为建立国家标准《信息处理用现代汉语分词词表》而制订的规范(草案),目的是广泛征求意见,引起讨论,以期改善之。

关键词: 中文信息处理, 汉语自动分词, 词表, 国家标准

The Guideline of Constructing a Wordlist of Contemporary Chinese for Information Processing

Sun Maosong¹ Wang Hongjun² Dong Xiufang²

National Key Lab. of Intelligent Technology and Systems, Tsinghua University¹

Department of Chinese Linguistics, Peking University²

E-mail: sms@s1000e.cs.tsinghua.edu.cn

Abstract: The guideline of constructing a wordlist of contemporary Chinese for information processing, expected to be a national standard of China, is presented. Suggestions are welcome so as to improve the quality of the guideline.

Keywords: Chinese information processing, Chinese word segmentation, wordlist, national standard of China.

国家标准《信息处理用现代汉语分词规范》对推动汉语自动分词的发展起到了积极作用[1]。然而,也有普遍反映,其中反复出现的表述“结合紧密、使用稳定(频繁)”,虽试图对某些介乎“词”与“词组”之间的“分词单位”进行界定,但这个表述很模糊,难以具体操作。来自语言学和计算语言学两个领域的专家们建议,应配一个词表作为该规范的补充,凡收进词表的“分词单位”,即认为是“结合紧密、使用稳定(频繁)”,否则不是。以此策略将该规范中存在的“灰色地带”明确地区分开来。我们接受了教育部语信司的任务,目标是建立这样一个分词词表,并最终成为国家标准(已在国家质量监督检验检疫总局立项)。以下全文给出我们为建立《信息处理用现代汉语分词词表》而制订的规范(草案)。希望能广泛征求意见,引起讨论,以期改进之。

1 主题内容与适用范围

1.1 主题内容

本规范构造的词表与 GB/T 13715-92 《信息处理用现代汉语分词规范》相配合,在大规模汉

* 本研究得到国家社科研究“九五”规划重大项目(97@YY001)和国家973项目(G1998030507)的资助。

语语料库的支持下，通过列举的方式，对 GB/T 13715-92 所定义的“分词单位”进行显式界定，以满足信息处理的需要。本词表应具备较强的通用性及覆盖能力，对汉语信息处理中的规范化以及各种汉语信息处理系统之间的兼容性有重要作用。

1.2 适用范围

本词表适用于汉语信息处理各领域，其它行业和学科可以参考使用。

汉语信息处理各领域可以根据其专门需要，进一步扩充本词表。

2 引用标准

信息处理用现代汉语分词规范 GB/T 13715-92。

3 术语

以下术语部分引自 GB/T 13715-92，部分为本规范最新定义。

3.1 汉语信息处理 用计算机对汉语的音、形、义等信息进行的处理。

3.2 词 最小的能独立运用的语言单位。

3.3 词组 由两个或两个以上的词，按一定的语法规则组成，表达一定意义的语言单位。

3.4 成分 词和词组的统称。

3.5 分词单位 汉语信息处理使用的、具有一定的词汇性质以及确定的语义或语法功能的成分。包括词和 GB/T 13715-92 所谓“结合紧密、使用稳定”的词组。

3.6 静态分词单位与动态分词单位 本规范将分词单位进一步区分为面向词表的静态分词单位与面向文本的动态分词单位两种类型。GB/T 13715-92 所定义的实际上是动态分词单位，本规范则定义静态分词单位。两类分词单位的内涵与外延大体一致，但在某些情况下有所差异。例如，汉族人名“周恩来”为静态分词单位，但非动态分词单位；人名简称“小李”为动态分词单位，但非静态分词单位。

3.7 分词词表 由具有一定使用频度的静态分词单位所构成的集合。

3.8 汉语语料库 能够体现现代汉语基本面貌的、大规模真实文本的集合。汉语语料库在本规范制订过程中的作用为控制静态分词单位的频度。

3.9 汉语分词 针对信息处理需要，根据 GB/T 13715-92，对汉语按分词词表进行词法分析的过程。

4 总则

4.1 确定分词单位的一般原则（为表述方便，下文出现的所有“分词单位”，均指“静态分词单位”）

4.1.1 单字词全部为分词单位。如：“花、打、大、一、在、了、吧”。

本规范认同以下各类汉字属于单字词范畴：① 公认或者争议不大的单字词（注：名词性单字成分加缀“子”后，如果可单用，也将该字算作单字词，如“桌、裙、裤”）；② 全部化学元素及常用有机化学物；③ 重要地名简称；④ 中国朝代；⑤ 中国 56 个民族。

姓氏不算作单字词。

4.1.2 属于封闭类的词（代词、数词、量词、介词、连词、助词、语气词、叹词、象声词）全部为分词单位。

4.1.3 直接包含至少一个非自由语素的成分为分词单位。如：“惋惜、学习、禽流感”。

4.1.4 多音节单义成分为分词单位。如：“徘徊、沙发、巧克力”。

4.1.5 有转义的成分为分词单位。如：“黑板、吃醋、炒鱿鱼、喝西北风”。

4.1.6 虽未发生转义，但“结合紧密、使用稳定”（通过汉语语料库的频度予以控制），在句法中起类似词的作用、一般不超过四字的词组为分词单位。如：“猪肉、小桥、手表厂、长途汽车、喝

酒、接到、湖边”。

4.1.7 习用语（包括成语、俗语、谚语、古诗句、格言等）为分词单位。如：“胸有成竹、欣欣向荣、不管三七二十一、行百里者半九十、柳暗花明又一村”。

4.1.8 能产的四字格，可作为分词单位。如：“AXBY”（“增收节支”），“半A半B”（“半明半暗”），“不A不B”（“不男不女”），“忽A忽B”（“忽东忽西”），“可A可B”（“可亲可敬”），“时A时B”（“时好时坏”），“随A随B”（“随叫随到”），“无A不B”（“无商不奸”），“一A一B”（“一蹦一跳”），“一A二B”（“一来二去”），“有A有B”（“有说有笑”）“又A又B”（“又哭又闹”），“A来B去”（“走来走去”）。

4.1.9 缩略语为分词单位。如：“环保、奥运会、零部件、大中小学”。

中间含顿号的，一律不收。如：“初、中、高级”“初、高中”“上、下课”“大、中型”。

4.1.10 由字母与汉字组成的词为分词单位。如：“B超、卡拉OK、IC卡”。

4.1.11 专名（所指唯一）为分词单位。如：“北京、泰山、上海市、珠穆朗玛峰、阿尔卑斯山脉”。

4.1.12 分词单位与缀的组合为分词单位。如：“革命性、非金属、反法西斯主义者”。

注：缀指意义较虚化、能产性强、构词中位置固定的语素。本规范定义的“缀”与语言学中的“词缀”范围不尽一致。缀表见附录。

4.1.13 由数词参与构成、原则上可无限类推的成分，数词小于等于三者，为分词单位；数词大于三者，一律为非分词单位。如：“大叔、二叔、三叔”“第一、第二、第三”为分词单位，“四叔、五叔、六叔”“第四、第五、第六”则为非分词单位。

4.1.14 分词单位加儿化音“儿”，如果在书面上“儿”必须出现，则为分词单位。如：“当儿、个儿”。否则为非分词单位。如：“花儿、悄悄儿、玩儿”。

4.2 构造分词词表的一般原则

4.2.1 任一分词单位，只有当其在汉语语料库中的使用频度足够高（即超过某阈值）时，才能被收入分词词表。即进入词表的分词单位应具有常用性。常用性与“结合紧密、使用稳定”在本质上是相通的。

4.2.2 任一非分词单位，其使用频度再高，也不能被收入分词词表。

4.2.3 本规范将使用频度的阈值分为“常用”、“十分常用”、“特别常用”三级，依次增加。一般来说，分词单位只需“常用”即可进入词表（如“猪肉”），但某些具有特殊结构的分词单位要进入词表，则要求其必须“十分常用”（如分词单位与缀的任意组合），甚至“特别常用”（如“单字名词+单纯方位词”组成的方位结构）。

5 细则

本节阐述词表收录条目的细则。注意，被收入词表的一定是分词单位，非分词单位则不可能进入词表。与 GB/T 13715-92 的体例相呼应，本节仍按照词类展开说明（注意，这里所谓“词类”，既针对词，也针对具有一定词汇性质的词组）。词类体系仍沿用 GB/T 13715-92 所定义的名词、动词、形容词、代词、数词、量词、副词、介词、连词、助词、语气词、叹词、象声词十三类。需要申明一点的是，本规范总则和细则未能顾及的任何成分，均默认为不可收入词表。

5.1 名词

5.1.1 普通名词

5.1.1.1 常用二、三字名词和名词性词组（主要为“名+名”、“区别词+名”、“动+名”“名+动”之类的定中结构）收入词表。如：“火车、牛肉、男人、冻肉、洗脸盆、胃溃疡”。

5.1.1.2 “形+名”定中结构中，常用的二、三字名词性词组，有限制地收入词表。主要包括：① 形容词反映的是名词所指事物的典型属性。如：“咸盐、白雪、蓝天、绿叶、白兔”；② 形容词具有分类作用而不是临时指别作用。如：“大桥、小桥、高个子、矮个子”；③ 形容词与名词的组合有一定象征意义。如：“红旗”。

5.1.1.3 不带“的”的定中结构中，常用的四字及四字以上名词和名词性词组（包括已进入通用领域的专业术语），有限制地收入词表。如：“市场经济、环境保护、八级工资制、长途汽车、市内电话、激光打印机”。

5.1.1.4 含助词“的”的名词性词组，原则上不收入词表。个别已成为习用语的除外，如：“掌柜的、相好的、真格的”。

5.1.1.5 分词单位与缀任意组合成的名词性成分，十分常用的收入词表。如：“科学家、物理学、非党员、马克思主义、围观者”。

5.1.1.6 除“人们”之外，名词性分词单位加表复数的“们”组成的成分，不收入词表。“哥儿们、爷儿们”不属此列（属习用语）。

5.1.1.7 单字名词重叠式只收“人人”。名词性重叠式“AABB”，常用的可收入词表，如：“方方面面、山山水水、瓶瓶罐罐”。

5.1.1.8 常用名词性缩略语收入词表。如：“副总、奥运会、五讲四美、二为方向、外交部长”。

5.1.1.9 常用的简称并列，以及常用的单字名词并列可收入词表。如：“干群、工农兵、门窗、柴米油盐”。

5.1.1.10 方位词全部收入词表。如：“上、左、东、外、东南、右上、前面、左边、以下、之前”。

① “单字名词+上/下/前/后/中/里/外/内/旁/边”组成的方位结构，特别常用的可收入词表，如：“空中、江边、屋里、墙上”。“双字或双字以上名词+单字方位词”组成的方位结构，无引申义的一律不收入词表，如：“桌子上、商店里、电影院旁”。常用且有引申义的可收入词表，如：“组织上、国际上”；② “东/南/西/北+单字名词”组成的名词性词组，常用的可收入词表，如：“东门、西城”。“东/南/西/北+双字或双字以上名词”组成的名词性词组，一律不收入词表；③ “名词+东/南/西/北”组成的名词性词组，除已成为某地区的专有名称外（如“江北、辽东”），其余一律不收。如“城东、路西”；④ 含双字方位词的方位结构一律不收。如：“床上面、屋里头”。

5.1.1.11 时间名词收入词表。如：“昨天、明年”。

① 一年的十二个月以及每周的七天，全部收入词表。如：“五月、星期一、星期天、礼拜日”；② “阿拉伯数字+年/月/时/分/秒”，一律不收入词表。如：“1988年、5月”；③ 由于干支组合成的纪年名称，常用的收入词表。如：“甲午、庚子”；④ “十二生肖+年”全部收入词表。如：“鼠年、牛年”；⑤ 由地支与“时”组合成的纪时名称全部收入词表。如：“子时、丑时”；⑥ 五个更次、初一至初十全部收入词表。如：“一更、五更、初一、初十”；⑦ “前、后、上、下、大前、大后”等与时间名词或时间量词的组合，常用的收入词表。如：“前天、后年、上星期、下月、大前天、大后天”。

5.1.2 专有名词

5.1.2.1 中国民族名、朝代名全部收入词表。如：“傣、傣族、秦、秦朝、秦代”。

5.1.2.2 著名的汉族人名，可收入词表。如：“毛泽东、欧阳修”。

5.1.2.3 著名的其它国家、民族的人名，可按其习惯形式收入词表。如：“牛顿、小林多喜二”。

5.1.2.4 “姓氏+职务、姓氏+‘某’、姓氏+‘氏’”一律不收。如：“张总、李科长、王某、刘氏”。

5.1.2.5 “老/小/大+单字姓氏”一律不收。如：“老张、小李、大王”。

5.1.2.6 著名的地名，无论是否包含通名部分，均可收入词表。如：“密西西比、密西西比河、天安门、长安街、广东省、中关村”。

5.1.2.7 语言及文字的名称，常用的可收入词表。如：“英语、中文”。

5.1.2.8 国家名及著名的机构或组织名，可收入词表。如：“中国、中华人民共和国、中国共产党、新华社”。但机构名中含“公司、集团”之类的通名，则一律不收。如：“海尔集团、微软公司”。其中的专名部分，十分常用的可收入词表。如：“松下、微软”。

5.1.2.9 其它著名的专名（如作品名、商标名、历史事件或政治运动等）可收入词表。如：“红楼梦、奔驰、中华烟、五四运动”。但商标名中含“牌、型、号”等，则一律不收。如：“永久牌、牡丹3型”。

5.1.2.10 著名的专名简称，可收入词表。如：“鞍钢、一汽大众”。

5.1.2.11 地名简称并列后接普通名词，常用的可收入词表。如：“川藏公路、京广线”。

多个地名的简称并列，如果表示一个更大的地域，常用的可收入词表，如：“京津、京津唐、江浙”。否则不收，如：“津沪、中美”。

5.2 动词

5.2.1 单字动词重叠式“AA”及相关的“A一A、A了A、A了一A、AA看”式一律不收入词表，如：“想想、想一想、想了想、想了一想、想想看”。一般单字动词重叠式“A不A”“A没A”一律不收入词表，如：“说不说、说没说”（“有没有、是不是”除外）。单字能愿动词重叠式“A不A”，特别常用的可收入词表，如：“能不能、要不要”。

5.2.2 二字动词重叠式“AAB”及相关的“A一AB、A了AB、A不AB”式一律不收入词表，如：“洗洗澡”“洗一洗澡、洗了洗澡、洗不洗澡”。二字动词重叠式“ABAB”一律不收入词表，如：“研究研究”。动词性重叠式“AABB”，常用的可收入词表，如：“拉拉扯扯、说说笑笑、走走停停”。

5.2.3 能愿动词全部收入词表。

能愿动词与动词组成的结构一律不收入词表，如：“该来、能吃、会跑”。单字否定词（“不、没、未”）与能愿动词组成的二字结构，特别常用的可收入词表，如：“不会、没能、未能”。

5.2.4 趋向动词全部收入词表。

5.2.5 动宾结构

5.2.5.1 动词和宾语的意义都比较抽象、扩展受限的常用二字动宾结构可收入词表，如：“引才、抗洪”。

5.2.5.2 动词与宾语的搭配十分受限的二字及三字动宾结构，常用的可收入词表，如：“瞪眼、跺脚、包饺子、擀面条”。

5.2.5.3 取多义动词非常用义的二字及三字动宾结构，常用的可收入词表，如：“拿分、打油、打吊针、做生日”。

5.2.5.4 只有加“着”才能成立的三字或四字动宾结构，常用的可收入词表，如：“板着脸、绷着脸、硬着头皮”。

5.2.5.5 宾语部分有形容词修饰，但不能单说或单说意义有变化的三字动宾结构，常用的可收入词表，如：“出远门、守活寡、伸懒腰、拖后腿”。

5.2.5.6 动词或宾语的意义较为具体且动词和宾语都使用常用义的二字及三字动宾结构原则上不收入词表，如：“买酒、拿枪、吃苹果”。

5.2.5.7 动宾结构内部还套有其它谓词性结构的不收入词表，如：“擦净脸、打死人、买到鱼”。

5.2.5.8 二字动词与二字或二字以上宾语组成的动宾结构一律不收入词表，如：“打扫房间、保卫祖国、防治艾滋病”。

5.2.6 动补结构

5.2.6.1 补语表示动作结果的二字动补结构，动词与补语之间语义关联较为紧密、具有一定可预期性的，常用的可收入词表，如：“打死、收到、拧干”。补语表示目标偏离的不收，如：“挖浅、买多”。

5.2.6.2 补语为单字趋向动词（“上、下、来、去、进、出、入、回、过、开、起”）的二字动补结构，十分常用的可收入词表。如：“走来、拿出”。补语为二字趋向动词的动补结构，一律不收，如：“走过来、跑出去”。

5.2.6.3 由具有虚化倾向的补语（“完、住、掉、好、起、成、为、到表动作有结果”）组成的二字动补结构，十分常用可收入词表，如：“跑掉、捆住、变为、收到”。

5.2.6.4 必须带“得”、“不”才能成立的三字动补结构，可收入词表，如：“来得及、来不及、经得起、经不起”。其余带“得”、“不”的三字及三字以上动补结构一律不收。

5.2.6.5 “动词+介词”组成的二字结构，如果语义上关联紧密、能分析为一个整体（如，表附着或存现义的动词与介词“在、于”，表位移的动词与介词“到、向”等），十分常用的可收入词表。如：“放在、生于、跑到、走向”。不属于以上情况的不收，如：“吃在北京”中的“吃在”。

“动词+介词”组成的三字结构，只收介词为“于”且十分常用的，如：“有利于、来源于”。

5.2.7 状中结构

5.2.7.1 状语为表方式、状态的常用二字状中结构，可收入词表，如：“手举、盒装、前伸、斜挎、疾跑、偷拍、骤降”。

5.2.7.2 三字状中结构不收，如：“胡乱说、直叫唤”。个别搭配十分受限的除外，如：“嗷嗷叫”。

5.2.7.3 状语表程度的状中结构一律不收，如：“很怕、挺担心、十分想念”。

5.2.7.4 由表示、时间、连接、范围等的副词与动词组成的状中结构不收，如：“不说、刚来、才去、虽爱、就走、都去、只拿”。

5.2.7.5 状语表否定的二字或三字状中结构，如果没有对应的肯定形式，可收入词表，如：“不便、不道德、不愈”。否则不收。

5.2.8 主谓结构

5.2.8.1 二字主谓结构，谓语为非自动词或形容词且包含至少一个非自由语素，常用的可收入词表。如：“身残、体弱、质次、价廉”。否则一律不收。

5.2.8.2 三字及三字以上的主谓结构原则上不收入词表，个别比较凝固的主谓结构除外，如：“雾蒙蒙、清风徐徐、湖光粼粼”。

5.2.9 并列结构

单字动词组成的并列结构，常用的可收入词表，如：“听说读写、吃穿住行、比学赶帮超”。

二字动词组成的并列结构一律不收，如：“宣传鼓动、调查研究”。

5.3 形容词

5.3.1 形容词重叠式“AA、ABAB”及相关的“A不A、A不AB”式一律不收，如“胖胖、雪白雪白、黑不黑、容不容易”。形容词性重叠式“AABB、ABB、AAB、A里AB”，常用的可收入词表，如“高高兴兴、红彤彤、冰冰凉、绿油油、马里马虎”。

5.3.2 两个反义的单字形容词并列且改变词性的，常用的可收入词表，如：“大小、长短、深浅”。

意义关联紧密的单字形容词组成的并列结构，常用的可收入词表，如：“黑瘦、麻辣、低矮”。

5.3.3 有关颜色的形容词及其相应的名词性成分，常用的可收入词表。如：“红、深红、橄榄绿、浅黄、暗灰、深红色、淡绿色”。

5.3.4 用作区别词的成分，十分常用的可收入词表，如：“无糖、纯天然、单门”。

5.3.5 副词与形容词的组合一律不收，如：“很棒、特惨、不红”。

5.4 代词

5.4.1 代词全部收入词表。

5.4.2 “各、每、全、本、某、该、此、其、贵、我”等后加单字名词性成分组成的结构，特别常用的可收入词表。名词性成分为非自由成分的，常用的可收入词表。如：“各国、每人、全军、本报、某事、该生、此事、其人其事、贵刊、我校”。

“这、那、哪”与单字名词性成分组成的结构，不收，如：“这人、那年、哪家”。

“这、那、哪”与量词组成的结构，除量词为“个、种”的外，一律不收。

5.5 数词

5.5.1 个位数、位数词全部收入词表。

5.5.2 个位数与位数词的任意组合，一律不收入词表。表概数的两个位数的组合除外，如：“千百、百十”。

5.5.3 数词组成的专名，常用的可收入词表，如：“五四、十一”。

5.5.4 “第、其、之”与后面的数词组成的表序数的结构，除“第一、第二、第三、其一、其二、其三、之一、之二、之三”外，一律不收入词表。

5.5.5 分数中的“分之”“百分之”“千分之”收入词表。

5.5.6 表概数的两个位数的组合，常用的可收入词表。如：“千百、百十、千万”。

表概数的两个数词的组合，一律不收入词表。如：“三四、三五、七八”。

5.5.7 “成、上”在位数词前，表示概数时，可收入词表，如：“成百、上千”。

“近、约、数、几”等在数词、位数词前表示概数时，则一律不收，如：“近千人、约三百、数万、几十”。

表概数的习用语，常用的收入词表，如：“百儿八十、万儿八千”。

5.5.8 “些、一些、点、一点”等表概数的词与形容词或动词组成的结构，一律不收入词表，如：“大些、懂一些、快点、快一点”。

5.6 量词

5.6.1 单字量词全部收入词表。

5.6.2 复合量词，常用的可收入词表，如：“人年、人次、架次、吨公里”。

5.6.3 数量结构（除“一个、一种、几个、几种”外），一律不收入词表。

5.6.4 量词重叠式“AA”（除“个个、种种、处处”外）一律不收入词表，如：“件件、张张”。

常用的“AABB”可收入词表，如：“家家户户、点点滴滴”。

5.7 副词

5.7.1 副词重叠式“AA、AABB”，常用的收入词表，如“刚刚、常常、仅仅、陆陆续续”。“ABAB”式重叠一律不收，如：“非常非常、如何如何”。

5.7.2 两个单字副词的关联紧密的组合，特别常用的可收入词表，如：“从未、不曾、也就”。

5.7.3 起副词作用的成分,十分常用的可收入词表,如:“越来越、时不时、不得不、不能不”。

5.8 介词

5.8.1 介词全部收入词表。

5.9 连词

5.9.1 连词全部收入词表。

5.10 助词

5.10.1 助词全部收入词表。

5.10.2 “动词/形容词+助词‘了/着/过’”一律不收入词表,如:“吃了、红了、走着、看过”。“掰了、罢了、除了、来着”等不在此列。

5.10.3 助词“所”与其后动词组成的结构,一律不收入词表,如:“所说,所想,所认识”。

5.10.4 助词“之”与其后单字组成的二字结构,特别常用的或其后单字为非自由成分的收入词表。如:“之最、之际、之辈、之滨、之巅”。

5.11 语气词

5.11.1 语气词全部收入词表。

5.12 叹词

5.12.1 叹词全部收入词表。

叹词重叠式“ABB”,常用的可收入词表。如:“哎呀呀”。

5.13 象声词

5.13.1 单字象声词,常用的收入词表,如“咚”。

5.13.2 “AA”式象声词,常用的可收入词表,如:“咚咚”。

5.13.3 单字象声词三字以上重叠一律不收,如:“咚咚咚、咚咚咚咚”。

5.13.4 象声词重叠式“AB、ABB、AABB、A里AB”,常用的收入词表,如:“噼啪、啪啦、啪啦啦、噼噼啪啪、噼里啪啦”。象声词重叠式“ABAB”一律不收,如“轰隆隆隆”。

[附录] 缀表(带*号的是学界比较承认的缀,不带*号的意义较实,因其高能产性而收录)

前缀:次,亚,代*,反,泛,副*,类,总*,半*,超*,纯*,防*,非*,抗*,可*,无*

后缀:部,厂,处(机构),度*,队,法*(法律名),费*,分子*,感,机*,家*,界*,局,科(机构、动植物类名),佬,类*,量*,论,率*,迷,品*,器*,区,热,人*,式*,厅(机构),系,形,型*,性*,学*,业*,员*,者*,制*,种,主义*,状*,族*,化*

在这个规范的指导下,我们已经初步建立了一个包含9万余条目的《信息处理用现代汉语分词词表》[2]。当然,要使这个词表上升为国家标准,还要经过一系列审慎的步骤。而我们在这里将词表规范(草案)公之于众,正属于其中一个重要环节。敬请专家学者不吝赐教。

参 考 文 献

[1] 国家技术监督局. 中华人民共和国国家标准 GB/T 13715-92《信息处理用现代汉语分词规范》. 中国标准出版社, 1993.

[2] 孙茂松、王洪君等. 信息处理用现代汉语分词词表. 语言文字应用, 2001, 4: 84-89.