

# 常用双音释词词量及提取方法

——对《现代汉语词典》双音同义释词的量化分析

苏新春 孙茂松

**提要** 一种语言中最重要的词语有多少,是认知科学、语义学、词典学中的重要命题。本文提取出《现代汉语词典》中单义的双音释词 6010 例,根据频次与释词位置进行加权,再用同释词、转释词、同素词三种方法来进行系联,提取出了 502 条常用释词。发现释词频率高,释词位置靠前的一般都具有通用性强、词义覆盖面广、语义位置重要的特点。

**关键词** 同义词;释义语言;词频;《现代汉语词典》

## 一 同义释词的范围与选择标准

汉语语义系统中的核心词语有哪些,这是当前汉语词汇学界、语义学界,也是中文信息处理界非常关注的一个问题。大家给予的定名有所不同,有“义原”说(董振东、董强 2000),“定义原语”说(张津、黄昌宁 1996),“元语言”说(李葆嘉 2001)。诸说各有所重,但所探讨对象的基本语义特征却大体相同,就是它们都处在语义系统的核心位置,义域范围大,覆盖面广,幅射其他词语的力量强等。拥有了这样一套词语,就具有了对整个语言的解释力,就能涵盖这一语言的所有词语。本文使用了“释义语言”的术语,不仅因为它利用的材料来自于词典释义,还因为它希望能用这套有限的词语来达到完整释义的目的。《现代汉语词典》是当代最有影响的一部语文词典,它对词语的精确释义长期以来受到人们的推崇,而且它的释义是在非自觉地使用有限释义语言的习惯下形成的,因此用它来作为本研究的材料就更显其价值。

为了在一开始就使这一探索性的工作更清晰些,我们先撇开词典中所有定义、说明式的释义,只探讨同义词对释的现象,并只限于双音单义词的同义词对释,以便更好地发现它们的语义特征与语义关系。《现代汉语词典》(1996)对同义词对释的运用有较严整的表述形式,如:

**【按脉】**诊脉。(按:无例句)

**【黯淡】**暗淡;色彩~。(按:释义与例句之间用“:”隔开)

**【弁言】**〈书〉序言;序文。(按:标示了〈方〉、〈书〉、(~儿)等词义特征)

**【充塞】**塞满;填满;库房里~着杂乱物品。(按:有两个释词,之间用“;”隔开)

以上是同义词对释的基本形式。有的是多种形式的交错使用,但只要具备了以上特点的都在本文的统计范围之列,如:

**【熬心】**〈方〉心里不舒畅;烦闷。(按:两个释义单位中只有一个双音释词)

**【充斥】**充满;塞满(含厌恶意);不能让低质量的商品~市场。(按:第二个双音释词使用了括号补充释义)

【奉达】敬辞,告诉;表达(多用于书信):特此~。(按:前面有词语属性的说明)

【风帆】船帆◇鼓起生活的~。(按:比喻性例句)

本文把同义释词界定为:它是一个具有较明确独立性的双音同义词释义单位。《现代汉语词典》中的双音同义释词在不同语境中,有着不同的诠释价值,与被释词的词义亲疏关系也不一样。用“外圆黑底阿拉伯数字”表示义项,义项与义项之间有着较明显的差异性,各自独立地存在;用“,”表示这几个释词解释的是同一个义项,它们之间的联系相当近,合起来成为一个释义单位,共同来说明被释词;用“;”表示这几个释词解释的也是同一个义项,但之间有着较明显的独立性,是分别、依次地说明被释词。第一种和第二种都不在本文的论述范围。本文分析的对象是《现代汉语词典》中的单义词,并有较明显独立性的双音释词。

我们先用 SQL 语言对《现代汉语词典》数据库进行条件设定,把所有符合以上释义特征的词语从数据库中调出,再进行人工干预,把“【口子<sup>1</sup>】量词,指人:你们家有几~”、“【绝早】极早:~动身”这样自由词组式的双音结构剔除在外,得出同义释词例 6010 条。《现代汉语词典》全书双音节词目 39800 余条,单一义项的 31500 余条,同义释词例约占 20%。

## 二 同义释词的统计与分析

《现汉》对一个单义词使用的释义单位最多有 4 个,如:

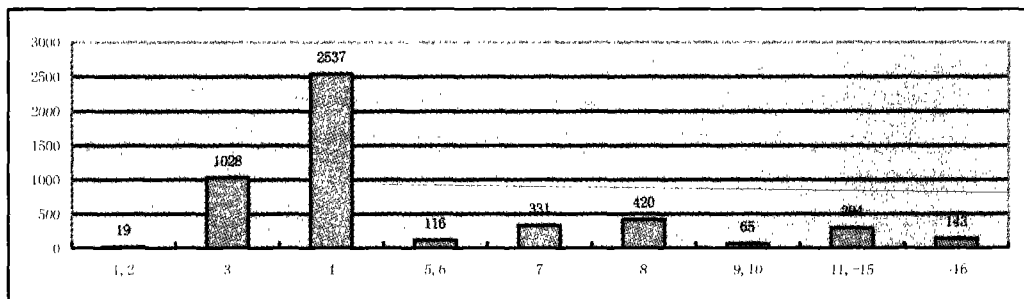
【便当】方便;顺手;简单;容易:这里乘车很~|东西不多,收拾起来很~。

这里使用了“方便”、“顺手”、“简单”、“容易”4 个释义单位。本文把第 1 个释义单位称之为“释词一”,其余依次为“释词二”、“释词三”、“释词四”。6010 个例词中所有四个释词位置的双音词语加起来一共是 7270 个(次),不重复的词语是 4953 个。

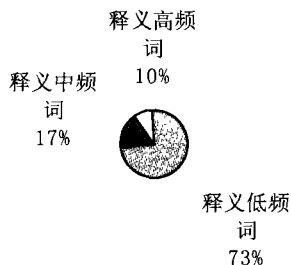
	释词一	释词二	释词三	释词四	总数
数字	5112	2124	32	2	7270
百分比	70.3%	29.2%	0.4%	0.02%	100%

处在不同位置的释词,它们在构成词语之间的同义关系时所起的作用是不同的。释词一显然要比位于后面的词语重要。愈是位于前面的词语,它们与被释词愈贴近,同义程度愈高。因此,在统计同义词的亲疏关系时,我们对不同释义位置的同义词采取了加权的方法,以便它们的词义关系能以适当的形式显示出来。加权为递减法,即只有一个释词的与位于释词一的均为 4 分,释词二为 3 分,释词三为 2 分,释词四为 1 分。这样把每一个释词处在不同位置上的得分相加,就得出这个词的权数。例如“聚集”,它位于释词一出现了 9 次,位于释词二出现了 3 次,这样它的得分就是  $4 \times 9 + 3 \times 3 = 45$  分。

4953 条同义释词的平均分值为 5.42。得分的数段分布如下:



示意图表明,得分最集中的在3至4分。4分的达2537,占到一半略多,其次是3分,为1028例,再加上1—2分的19例,合起来共有3584,占总数的73%。它们基本上属于释词一或释词二,且只释词一次。这样低分值、低频率使用的释词对说明同义词之间的类聚关系价值不太大。而5至8的得分,显示它们充当释词的次数基本上在2次。这样的词共有867个词,占总数的17%。值得注意的当然是释词出现次数多、分值高的词语,以9分(包括9分)来划界的话,在这个范围的词语有502条,占总数的10%,它们充当释词的次数起码在3次以上。这三部分词语可依次称为“释义低频词”、“释义中频词”、“释义高频词”,图示如下:



对本研究来说,最有价值的是只占10%的那502条,从宽的话则可把分值在5—8之间的867条词包括进来。对词典释义语言常用词量的提取来说,这是很重要的一份材料。为了有效地对比,下面的统计是将三种频度的释词放在一起进行,以便清楚地看到不同频度词语的状况。

#### 1. 释词充当词目的调查。

作为《现汉》词目		
释义低频词	3181条	88.7%
释义中频词	816条	94.1%
释义高频词	482条	96%

释词是来自于词典的释义语言,而观察这些释词是否又充当了《现汉》的词目,可以达到两个目的,一是看这些释词的稳定性如何,二是可以看到同义词之间的互释情况。从上面的数据可以看到,这三部分释词总的来说稳定性都相当高,但仍有所差异,即同义释词出现的次数愈多,表明它充当“词”的机率也就愈高。没有充当词目的大都为以下几种情况:(1)释词归入了单字头,如“荸荠、蹦子、鲑鱼、鲟鱼、柚子、蟋蟀、啜嚼、蛻螂、猓猓”等;(2)明显是词组的,如“被骗、办完、奔向、变胖、变瘦、别想、别动、含有、冲出、出声、云里、搓澡、打死、摘取、摘去”等;(3)构词词素浅显的,如“纯金、高楼、公马、公鸡、公驴、公牛、鸡蛋、活猪”等。

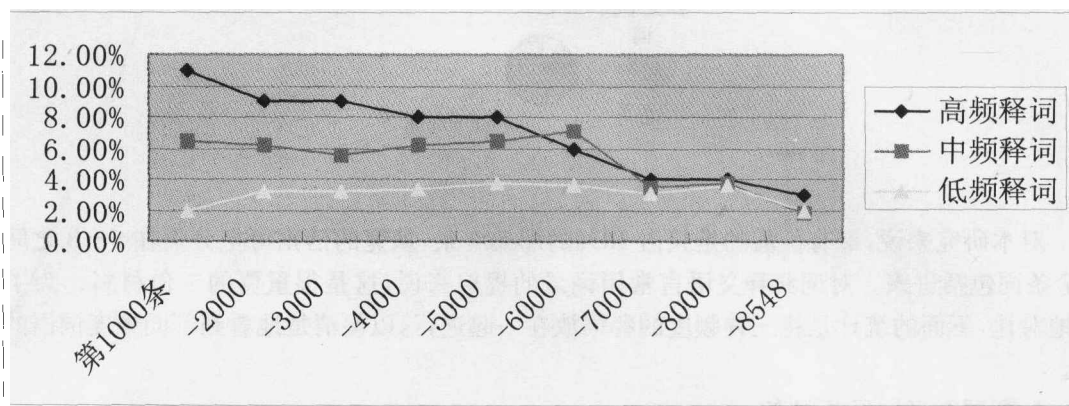
#### 2. 词语常用度的调查。

在《频率词典》中出现		
释义低频词	1013条	28.2%
释义中频词	416条	47.9%
释义高频词	322条	64.1%

这项调查使用的材料是《现代汉语频率词典》中的8458条常用词。这部词典对词语的切分与常用性的认定,带有中文信息处理的性质(苏新春1987),但由于它调查语料量适中,统计方法科学,所得出的数据相当可信。本调查可以观察到充当释词的机率高与使用频率高低

之间的关系。通过统计数字的对比,可以清楚地看到高、中、低频的释词在《频率词典》中所占的比例有着明显的递减趋势。由 64.1%到 47.9%,再到 28.2%,出现了三个明显的递减阶梯。

下面再看高频释词、中频释词、低频释词在《频率词典》8548 条中每千条的分布情况,第 1 条是使用度排名最前的,第 8548 条是排名最后的。数据显示,三种释词在使用度高低不同的词中逐段分布的差异也很明显。高频释词有 11%的词出现在前 1000 词中,而低频释词只有 1.9%。这种差距直到第 7500 条左右才逐渐消失。这说明在《现汉》中释词使用得愈多的词语,在《频率词典》的常用词中占的比例也愈高,位置也愈靠前。二者使用的方法虽然不一样,却显示出相同的词语规律,在其中起到了相互印证的作用。



### 3. 语法属性的调查。

这项调查反映释词语法属性的分布,可以观察到充当释词的机率高低与语法属性分布有何种联系。下面只统计位于释词一位置上双音同义释词的词性,不算重复的有 3779 个词,按重复算是 5112 词次。5112 词次的词性分布如下:

	动	名	形	副	连	其他
5112 词次	2454	1790	556	239	68	5

这项调查还做得比较粗糙,如只统计了释词一的词性,而未能统计所有释词的词性,而更多的是受到语料的限制,如释词的词性模糊,有两可的现象,“【依据】根据”可作动词也可作介词,“【回佣】回扣”可作动词也可作名词;又如有的释词兼有两种词性,“【为伍】同伙;做伙伴:羞与~”;再如有的是例句词性的并有,“【援助】支援;帮助:国际~|经济~|~受难者”,“【祝贺】庆贺:~你们超额完成了计划|向会议表示热烈的~”,均为名动两用。但这个初步调查还是有价值的,它清楚地显示了使用同义词释义的主要所属词类,在数量的多少上依次表现为“动→名→形→副→连”。

### 三 同义释词的语义特征与义类归纳

同义释词起着其他词的词义参照物的作用,充当同义释词的次数愈多,显示它们关涉到的词义对象也就愈多,适用面愈广,常用度愈高。词汇学理论上有一条重要规律:“词的意义范围是以它的同义词(包括文体同义词和意念同义词)为界限的。不知道一个词的同义词及其相互界限就很难掌握一个词的意义”(李友鸿 1958)。根据充当同义释词的多寡可以判断一个词的义域广狭。当一个词较经常地被用来作为其他词的同义代替物出现时,其实它也就取得

了这类同义词的代表者的身份,即这类词语的主题词、核心词。

“它的周围聚集了同义词群的其他成员,它的意义对于其他成员都是共同的,……核心词的重要作用,是联系本组同义词成员的纽带,是限定本组同义词范围的尺度。加入同义词群的每一个成员,都必须跟其核心词(或其中的一个意义)具有同义关系,否则便不能加入。这样,同义词群不仅有了核心,而且有了范围,就会形成一个封闭式的系统”(张志毅 1980)。同义词释义是由已知到未知,由熟知的词语来把握未熟知的词语,在它的背后,能够清楚地再现出同义词之间的语义联系,反映出同义词群中首词的语义特点与语义作用。本文对《现汉》同义词群进行系联时,使用了三种方法:

1. 同释词。指许多词用了一个共同的释词,这个释词往往就是同类词语中的主题词、核心词、广义词。下面试以两组词为例:

第一组:“聚集:屯聚、汇合、汇拢、汇集、荟萃、骈阗、凑集、萃聚、集拢、聚拢、凝集、麇集”。“聚集”解释了 12 个词,具体显示如下:

【屯聚】聚集(人马等):~大量兵力。

【汇合】(水流)聚集:会合:小河~成大河◇人民的意志~成一支巨大的力量。

【汇拢】聚集;聚合:几股人群~在一起|~群众的意见。

【汇集】聚集:~材料|把资料~在一起研究|游行队伍从大街小巷~到天安门。也作会集。

【荟萃】(英俊的人物或精美的东西)会集;聚集:~一堂|人才~。

【骈阗】<书>聚集;罗列;众多:士女~。也作骈填、骈田。

【凑集】凑在一起;聚集:人烟~|~技术力量。

【萃聚】<书>聚集:群英~。

【集拢】聚集:场院中~了一群人。

【聚拢】聚集。

【凝集】凝结在一起;聚集:心中疑云~|诗篇~着诗人对祖国的真挚感情。

【麇集】<书>聚集;群集。

由此可以概括出“聚集”是这一组同义词的主题词,它具有词义宽泛、适用对象多、涵盖面广、使用频度高、常用性强的特点。而其他 12 词在词义上都有或雅俗,或狭小,或僻用的特点,用“聚集”来作它们的释词,显义作用相当明显。而例句的出示则有效地克服了用广义词释狭义的种种不足。如中性的表义“屯聚”、“集拢”、“聚拢”、“汇合”,褒扬性的表义“荟萃”、“荟萃”,只与思维、心理、感情等搭配的“凝集”,既可与人也可与物搭配的“汇合”、“集拢”、“聚拢”等。这些词的例句凸现了该词的词义特点,也恰好反过来证明了这些词语词义范围狭窄的特点。

把“聚集”组词语再放到《同义词词林》中来考察,它们的分布情况是这样:

Ie10-01 屯聚、聚集、麇集、聚拢

Ie10-03 麇集

Ie10-05 荟萃

Ie08-04 汇合

《同义词词林》没有收录的词有“汇拢、汇集、骈阗、凑集、萃聚、集拢、凝集”,但收了单字“聚”、“集”、“凑”、“汇”,这四个单字都在 Ie10-01。由此可见,除了“汇合”一词外,其他 12 个词都在第一层 I 类“现象与状态”、第二层 e 类“事态”、第三层 10 类“集合 分散”之中,而且大部分又在“01”小类。《现汉》中通过同义词释义反映出来的词群与《同义词词林》的义类分布有着

相当贴切的对应关系。《同义词词林》是一部语义分类词典，它建立在同义词的基础上，但又比同义词的范围要宽泛得多。

“从总的方面来说，词群的组成与同义词的组成都必须以词的基本意义相同为原则，这是它们相同的部分；但是，词群的首词所带领的一群词比同义词的范围要广泛得多，这是由于词群是以概念为单位，而同义词是以词为单位的，这是它们最根本的不同。……词义分类词典的词群大于词汇学中的同义词，它除了包括严格的同义词外，还包括一个上位概念所属的下位概念的词”（鲍克怡 1985）。

可《现汉》中表现出来的首词是“聚集”，《同义词词林》的首词是“集合”，这里的原因是什么呢？可能与语义差异和语体差异都有关联。一是《同义词词林》的分类选词更注重适用，更贴近口语化，二是它的分类是“同”与“异”、“正”与“反”的词语同处一类，而“集合”与“分散”形成更工整的对，所以 Ie10 的类名用了“集合”而不是“聚集”。

第二组：“如果：假如、假若、假使、如其、如若、若是、设使、向使、要是”。“如果”解释了 9 个词，具体显示如下：

【假如】如果：～明天不下雨，我一定去。

【假若】如果：～你遇见这种事，你该怎么办？

【假使】如果：～你同意，我们明天一清早就出发。

【如其】如果。

【如若】如果。

【若是】如果；如果是：他～不来，咱们就找他去|我～他，决不会那么办。

【设使】假使；如果。

【向使】<书>如果；假使。

【要是】如果；如果是：～你想参加，我可以当介绍人|这事～叫他知道了，一定会发生争吵。

把这一组词放到《同义词词林》考察，发现这 10 个词全部属于 Kc08—01 类，即第一层 K 类“助语”、第二层 c 类“联接”、第三层 08 类“如果 那末”、第四层 01 小类。

现在来得出下面的结论就不难了，即《现汉》用同义词释义串联起来的词群就是一个小的义类，经常充当释词用的词就是这个义类的词首词、代表词，而 502 个高频同义释词就代表着 502 个小义类，这些高频词应该就是词典释义中最经常用到的常用词。由被释词的多少来透视释词义域范围广狭的方法，以前苏新春（1987，1990）在讨论古汉语基本词汇的时候也尝试过，这个方法相当可靠。

2. 转释词。指的是相互解释的词。可以通过转释词来找到词与词之间的联系，从而扩大词群的联系范围，把一个个孤立的义类串联成一个大的义类。

仍以“聚集”组为例。“聚集”是由“集合”来解释的，而“集合”又解释了“集聚、纠合、聚齐”；再下去，“纠合”又解释了“纠集”，“聚齐”解释了“会齐”。细列如下：

【聚集】集合；凑在一起：～力量|～资金|广场上～了很多人。

【集聚】集合；聚合：人们～在老槐树下休息。

【纠合】集合；联合（多用于贬义）：～党羽，图谋不轨。也作鸠合。

【聚齐】（在约定地点）集合：参观的人八时在展览馆门口～。

【纠集】纠合（含贬义）。也作鸠集。

【会齐】聚齐：各村参加集训的民兵后天到县里～。

这样通过互训串联,“集合、集聚、纠合、聚齐、纠集、会齐”就与前面的词合起来成为有 19 个词的同义词群。而这 6 个词在《同义词词林》中的分布也十分贴近,“集合、集聚”属 Ie10—01 类,“聚齐、会齐”属 Id10—02 类,“纠合、纠集”属 Ie10—04 类。

如果只停留在一个层次的同义释词,“聚集”、“集合”、“纠合”、“聚齐”就成了四个不关联的释词,它们的加权得分分别是 45、16、4、4。

3. 同素词。通过查找相同词素的方法来系联同义词词群。汉语复音词大部分是复合词,复合词的词义大都保留了词素的意义要素,因此,同素词群也就因为有同一词素而具有了某种程度的相同词义成分的可能。

上述“聚集”群,在 19 个词中共使用了 16 个词素,它们是“集 9”、“聚 6”、“拢 3”、“汇 3”、“合 3”、“纠 2”、“萃 2”、“齐 2”、“会 1”、“屯 1”、“骈 1”、“麇 1”、“凑 1”、“阨 1”、“凝 1”、“荟 1”,汉字后面数字表示构词数。在 5048 个同义释词中,用这 16 个词素构成的词有 82 个,其中整个词的词义属于“聚集、集合”类而又没有被上面的系联法所覆盖的词还有 5 个词。下面是这 5 个词的加权得分、《同义词词林》中的义类分布及所解释的词语:

集中:4 分/Ie10—01/集注

会合:10 分/Ie10—01/汇合、交会、交汇

聚合:6 分/Ie10—01/汇拢、集聚

群集:3 分/Ie10—01/麇集

会集:4 分/无/荟萃

通过以上同释词、转释词、同素词的三种系联方法,在 4953 个双音释词中就串联起了同属“聚集”义类的 8 个同义释词,其中属于高频释词范围的有 3 个,中频的 1 个,低频的 4 个。这就给最终提取该义类的代表词、核心词提供了扎实的基础。这仅是一项示范性的工作,它显示提取、概括词群的代表词、核心词的工作是可行的。这首先是因为这样的词语是客观存在的。同属一个词群中的“同义词有相当大的一部分没有直接的、自由的指名意义,这种同义词不是直接表示自己的基本意义,而是通过语义上的基本词或称核心词来表示的,这核心词的指名意义才直接通向现实,它是一连串相关同义词的基础”(苏·维诺格拉多夫 1958)。同义词同类词的核心词总是会以种种方式体现出来,它存在于人们的语言使用习惯,存在于种种语言材料,并总会在语用中首先受到人们的关注,成为常用词、高频词。这在成系统、大词语量、释义规范的语文词典中,会表现得尤为鲜明。

本文所作的只是提取词典释义常用词词量的初步工作。后续工作可以从多方面深入,如对高频与中频同义释词进行全面的义类概括;把分析对象从“单义”、“双音词”扩大到多义词,扩大到单音词与双音节以上的复音词;特别是对使用了说明式、定义式释义词例的分析,从对释式的常用词量延伸到定义式的常用词量的提取。目前所作的初步工作已显示了这一研究的诱人前景。它不仅将形成提取常用释义词词量的方法,还将揭示汉语词汇中最富于阐释力、最具覆盖力的核心词语系统,这类常用词量将为语言教学、教材编写、词典编纂、中文信息处理等领域提供进一步开发的基础成果。

#### 参考文献

鲍克怡 1985 试论词义分类词典的编纂,载《词典和词典编纂的学问》,上海辞书出版社。

董振东、董强 2000 《关于知网——中文信息结构库》(知网主页 <http://www.keenage.com>)。

- 李葆嘉 2001 现代汉语元语言系统研究和汉语词汇规范,全国汉语词汇规范问题学术研讨会(厦门)论文。
- 李友鸿 1958 词义研究的一些问题,《西方语文》第1期。
- 苏新春 1987 论古汉语基本词汇的广义性,《广州师范学院学报》第1期。
- 苏新春、顾江萍 1990 如何确定古汉语基本词汇的广义性,《广州师范学院学报》第1期。
- 苏新春、顾江萍 2001 “人”“机”分词差异及规范词典的收词依据——对645条常用词未见于《现汉》的思考,《辞书研究》第2期。
- 苏·维诺格拉多夫 1958 词的词汇意义的主要类型,《俄语教学与研究》第2期。
- 张津、黄昌宁 1996 《从单语词典中获取定义原语方法的研究及现代汉语定义原语的获取》,国家自然科学基金重点基金项目。
- 张志毅 1980 同义词词典编纂法的几个问题,《中国语文》第5期。

## The Number and Extraction Methods of Commonly Used Disyllabic Defining Words

### ——An Quantitative Analysis on the Disyllabic Synonymous Defining Words in *Modern Chinese Dictionary*

**Abstract** The number of the most important words in a language is a topic of great significance in cognitive science, semantics, and lexicography. From *Modern Chinese Dictionary*, the present authors first extract 6010 disyllabic defining words with single clearly specified meanings. By connecting these words with common defining representatives, mutual glosses and identical morphemes, and by weighing their importance in terms of frequency and position in definition, the authors further extract 502 commonly used disyllabic defining words. Finally they conclude that, generally speaking, the words that enjoy a high frequency in definition are characterized by a high currency in usage, a wide semantic scope and an important position in semantic meaning.

**Key words** synonyms; defining words; word frequency; *Modern Chinese Dictionary*

(苏新春 361005 厦门大学中文系)  
(孙茂松 361005 厦门大学计算机系)  
(责任编辑 张燕来)