



2007年 搜索质量盲测评估报告

北京正望咨询有限公司

China IntelliConsulting Corporation

吕伯望

peter@iaskchina.com

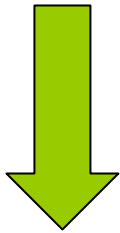
2007年6月



盲测设计

搜索质量盲测评估项目组成部分

盲测者招募



盲测前问卷
(个人信息)



盲测

(质量评估)



盲测后问卷

(使用习惯与态度)

搜索质量盲测评估项目组成部分

- 盲测者（样本）招募
- 盲测前问卷
- 盲测页面（关键词类别选定、关键词自由输入、判断、打分、说明理由）
- 每个盲测者的测试次数应做一定的限定
- 盲测结束后，告知盲测者分别判断百度好与Google好的次数，并询问是否符合其盲测前的预期
- 邀请盲测者继续填写一份更为深入详细的调查问卷

盲测系统设计原则

- 使用盲测前问卷获得盲测的用户特征分布情况
- 用脚本抓取不同搜索引擎的搜索结果并呈现在同一个网页上，提交给盲测者判断
- 不同搜索引擎的搜索结果随机出现在盲测页面的左/右区域
- 去掉搜索结果页面的Logo及其他特征文字或图案，同时尽量保持搜索结果的完整性和真实的原始面貌
- 算法设计保持严格中立，避免对被测搜索引擎产生任何主观偏向
- 由盲测者自主选取关键词，避免因关键词的限定设置导致任何可能的倾向性引导
- 通过盲测后一份更详细的在线调查问卷获得盲测用户的搜索行为与态度的统计数据，分析盲测结果与用户特征、用户行为及态度之间的关系

盲测项目认证书

清华大学



Tsinghua University

中国 北京 100084

Beijing 100084, China

北京正望咨询有限公司“搜索质量评估盲测”项目认证书

清华大学智能技术与系统国家重点实验室受北京正望咨询有限公司委托，对其“搜索质量评估盲测”项目所使用的软件系统进行技术认证。

经过对系统的文档和代码（包括 java 脚本和 html 页面）进行分析，并通过项目实施的两个星期内（4 月 9 日至 20 日）的在线试用与监测，得出如下结论：

1、该搜索引擎盲测项目的系统软件具有

（1）客观性。系统正确地抽取并完整地保留了被测搜索引擎的原始搜索结果，且未引入原始结果之外的其他内容，并如实显示在用户操作界面中；

（2）公正性。对参加盲测的各搜索引擎采用完全相同的处理方法，且算法设计中未引入对被测搜索引擎的主观偏向。系统以随机的方式选择搜索引擎的搜索结果所呈现的左/右区域，避免暗示用户所评价的结果出处。

2、在盲测过程中，由用户自主选取关键词进行测试，避免了对用户测试内容的倾向性引导，保证了盲测结果的公证性与客观性。

清华大学智能技术与系统国家重点实验室

2007 年 4 月 20 日



盲测过程

- 用户输入关键词
- 选择该关键词所属的类别（由用户自主判断）
- 用该关键词搜索（由盲测系统实时完成）
- 查看、对比两个搜索引擎呈现的不同的搜索结果
- 作出比较性判断（“哪个更好？”）
- 选择自己判断的理由
- 系统限定每个用户最少测试4次，最多8次
- 结束盲测后，选择盲测结果是否符合预期

招募原则

- 在互联网用户到达率最高或用户特征比较特殊的网站上投放招募广告（百度、谷歌/Google、雅虎中国及其他利益相关网站除外）
- 投放的网站及网站内的频道涵盖新闻媒体、IT/科技、财经、游戏娱乐、社区与网站站长联盟
- 招募广告保持中性原则，不渲染、不抽奖，避免一切可能的诱导

盲测者招募网站

- 招募广告（文字链接或条幅）于2007年4月9日至20日放置在如下主要网站：

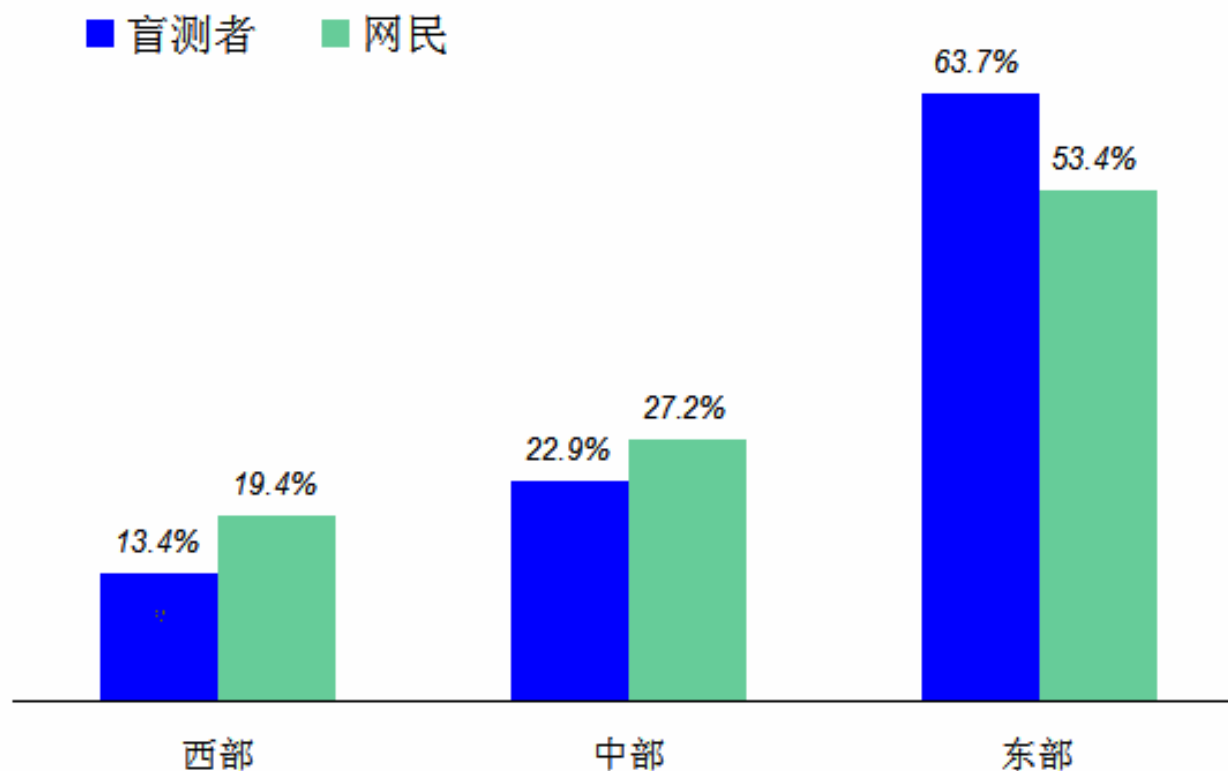


- 实到盲测者人数：2769人，其中中国大陆境内2720人，港澳台及海外合计49人



盲测者用户特征

盲测者的地区分布（1）



盲测者的地区分布（2）

前10省份	盲测者比例
北京	17.3%
广东	10.4%
上海	8.9%
浙江	7.0%
江苏	6.1%
湖北	4.9%
福建	4.7%
四川	4.5%
湖南	4.3%
山东	4.3%

前10城市	盲测者比例
北京	17.3%
上海	8.9%
广州	3.5%
武汉	3.2%
深圳	3.0%
成都	2.3%
杭州	2.1%
长沙	2.1%
天津	1.9%
南京	1.8%

盲测者的年龄与学历分布

年龄分布	盲测者比例	网民比例
18岁以下	3.2%	17.2%
18岁-24岁	39.9%	35.2%
25-30岁	38.5%	19.7%
31-40岁	16.9%	18.6%
41岁以上	1.6%	9.3%
合计	100.0%	100.0%

学历分布	盲测者比例	网民比例
初中及以下	3.1%	17.1%
高中（中专）	13.3%	31.1%
专科	22.3%	23.3%
大学本科	47.4%	25.8%
硕博士	13.9%	2.7%
合计	100.0%	100.0%

	盲测者	网民
学生比例	30.6%	32.3%

盲测者的上网与使用搜索频度

最近一次上网时间	
今天早些时候	60.6%
昨天	34.7%
一周以内	3.2%
一个月以内	0.3%
超过一个月 或更早	1.1%
合计	100.0%

每周上网几天	
天天上网	83.8%
5天以上	8.6%
3到5天	4.8%
1到2天	2.1%
每周不到1次	0.8%
合计	100.0%

使用搜索频率	
很频繁	56.2%
比较频繁	28.1%
一般	13.7%
偶尔用用	1.6%
今天以前 还没用过	0.4%
合计	100.0%



盲测结果

盲测结果概要

- 所有2769个盲测者共实测了11,864次，平均每人实测4.28次；实测4次的盲测者占了86.9%
- 在所有的11,864次实测中，判断Google搜索结果好于百度的比例48.2%，百度搜索结果好于Google的比例为39.8%，其余12.1%的测试，百度与Google一样好
- 在这次盲测设置的八个分类中，Google在7个分类中领先，只在娱乐搜索分类中输给百度5个百分点
- Google领先百度超过10个百分点的分类有：交通旅游（18.2个百分点）、购物餐饮（16.3）、互联网与IT（15.6）、以及教育科学文化百科（12.5）

盲测者完成的搜索测试次数

盲测中完成的搜索测试次数	占全体盲测者比例数
4	86.9%
5	6.5%
6	1.4%
7	0.9%
8	4.3%
合计	100.0%

我们的盲测系统限定每人至少测试四次，至多测试8次

盲测者的IP地址

单个IP地址提交的盲测结果数	IP地址数	提交结果数占全部盲测结果数的比例
14	1	0.5%
8	1	0.3%
7	1	0.3%
6	1	0.2%
5	3	0.6%
4	5	0.7%
3	6	0.7%
2	34	2.5%
1	2564	94.3%
合计	2720	100.0%

盲测关键词类别（1）

关键词类别	盲测次数	
新闻时事	1,695	14.3%
娱乐	2,713	22.9%
IT互联网	2,398	20.2%
财经工商	819	6.9%
购物、餐饮与生活服务	871	7.3%
交通旅游	1,237	10.4%
教育科学文化百科	1,596	13.5%
政务与公众信息	535	4.5%
合计	11,864	100.0%

随机从盲测者输入的11,864个关键词中选取100个进行分析判断，我们发现盲测者对关键词归类的准确性高达89.7%

盲测关键词类别 (2)

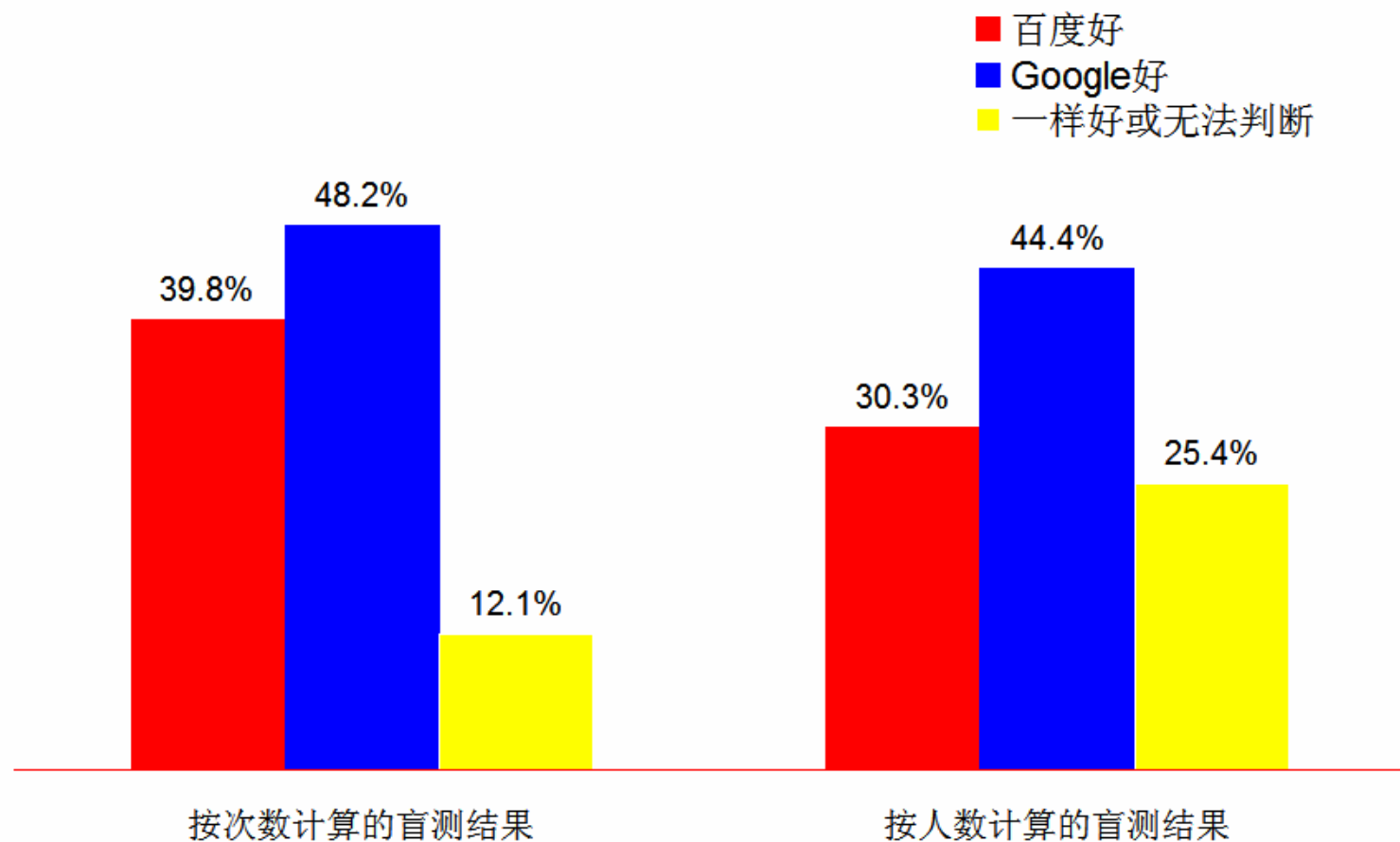
关键词类别	盲测次数			
	学生	非学生	高端用户	总体
新闻时事	14.5%	14.2%	14.0%	14.3%
娱乐	27.0%	21.1%	15.4%	22.9%
IT互联网	18.7%	20.9%	21.3%	20.2%
财经工商	4.8%	7.8%	9.3%	6.9%
购物、餐饮与生活服务	5.2%	8.3%	9.2%	7.3%
交通旅游	9.7%	10.7%	11.9%	10.4%
教育科学文化百科	15.8%	12.4%	13.7%	13.5%
政务与公众信息	4.2%	4.6%	5.3%	4.5%
合计	100.0%	100.0%	100.0%	100.0%

盲测总体结果

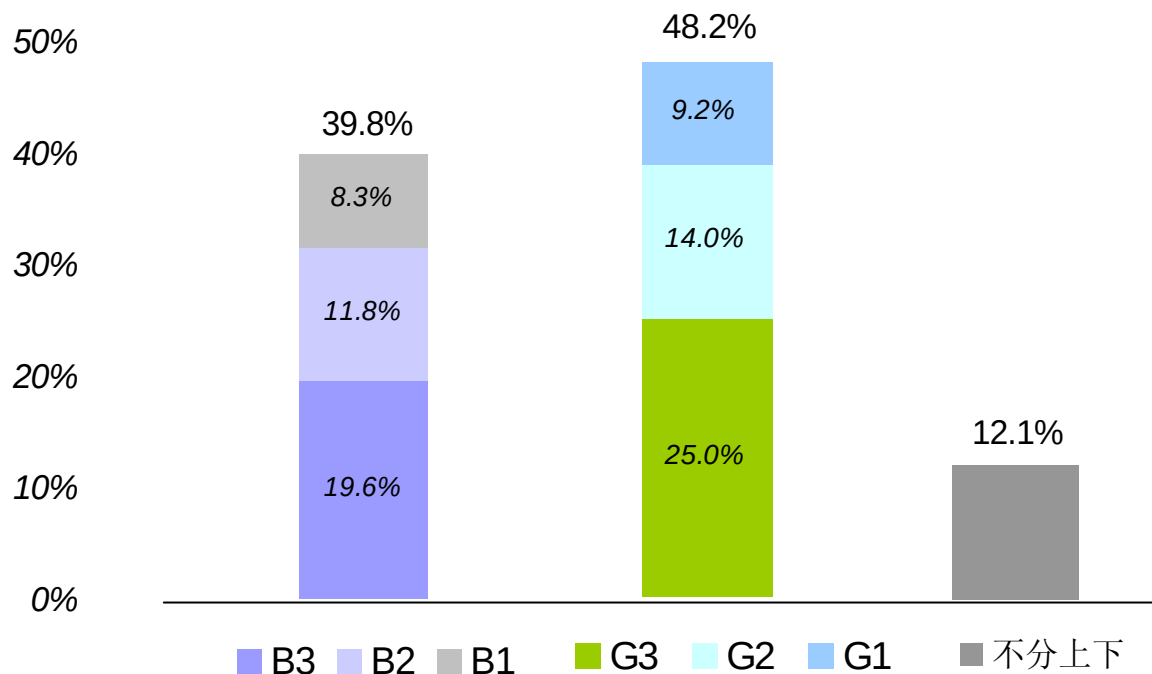
- 在所有测试的**11,864**次中：
 - 选择百度好的次数：39.8%
 - 选择一样好的次数：12.1%
 - 选择Google好的次数：48.2%
- 在大陆境内**2,720**人中：
 - 选择百度好的人数：30.3%
 - 选择一样好的人数：25.4%
 - 选择Google好的人数：44.4%

注：同一个盲测者测试了多次；如果他选择判断百度好的次数多于选择Google好的次数，则将此盲测者算作选择百度好的人

盲测总体结果



以实测次数计算的盲测结果——总结结果

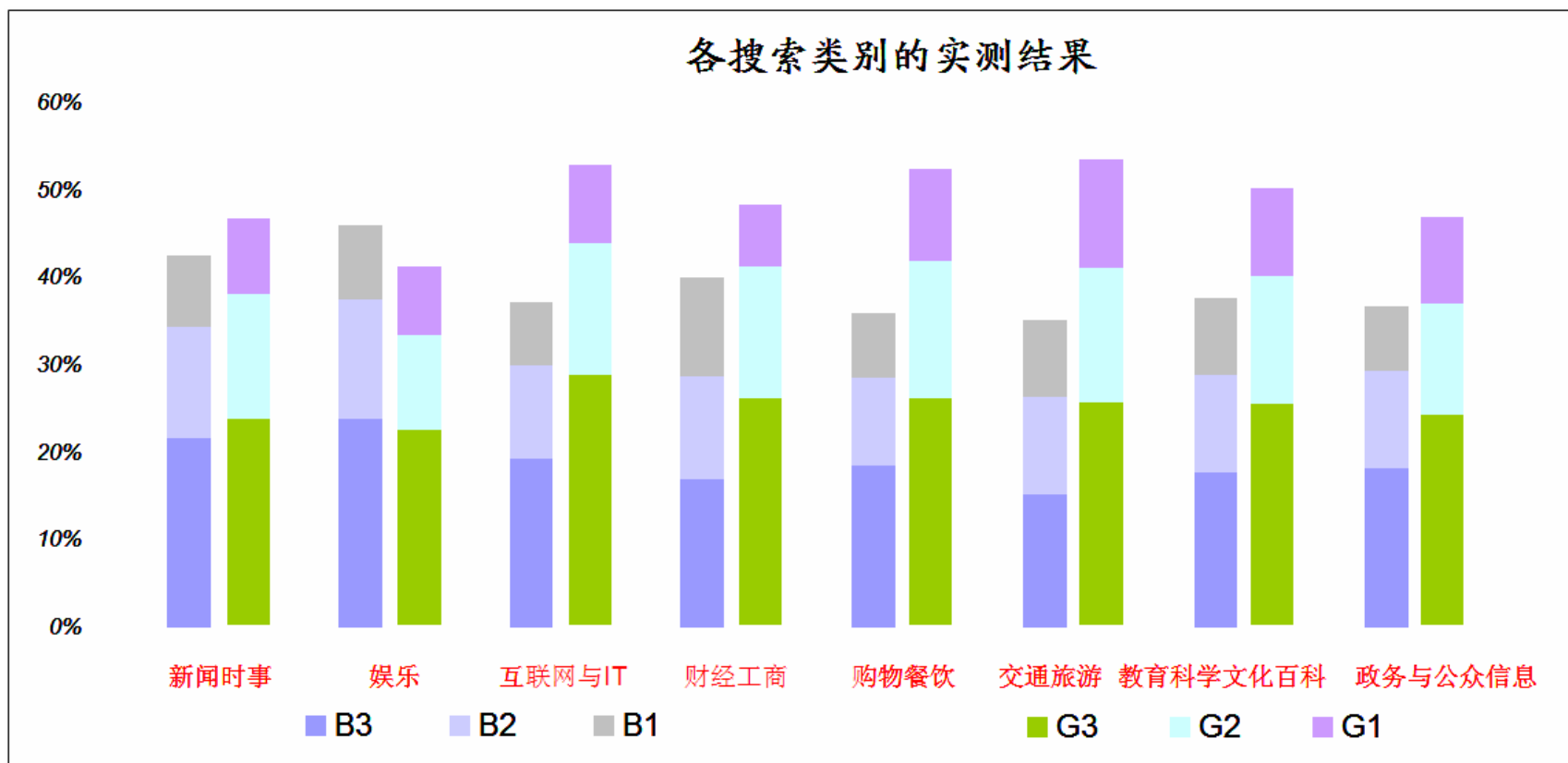


B3、B2、B1： 百度比Google好，给百度搜索结果分别打3分、2分和1分；

G3、G2、G1： Google比百度好，给Google搜索结果分别打3分、2分和1分；

3分： 好很多； 1分： 只好一点

以实测次数计算的盲测结果——分类结果

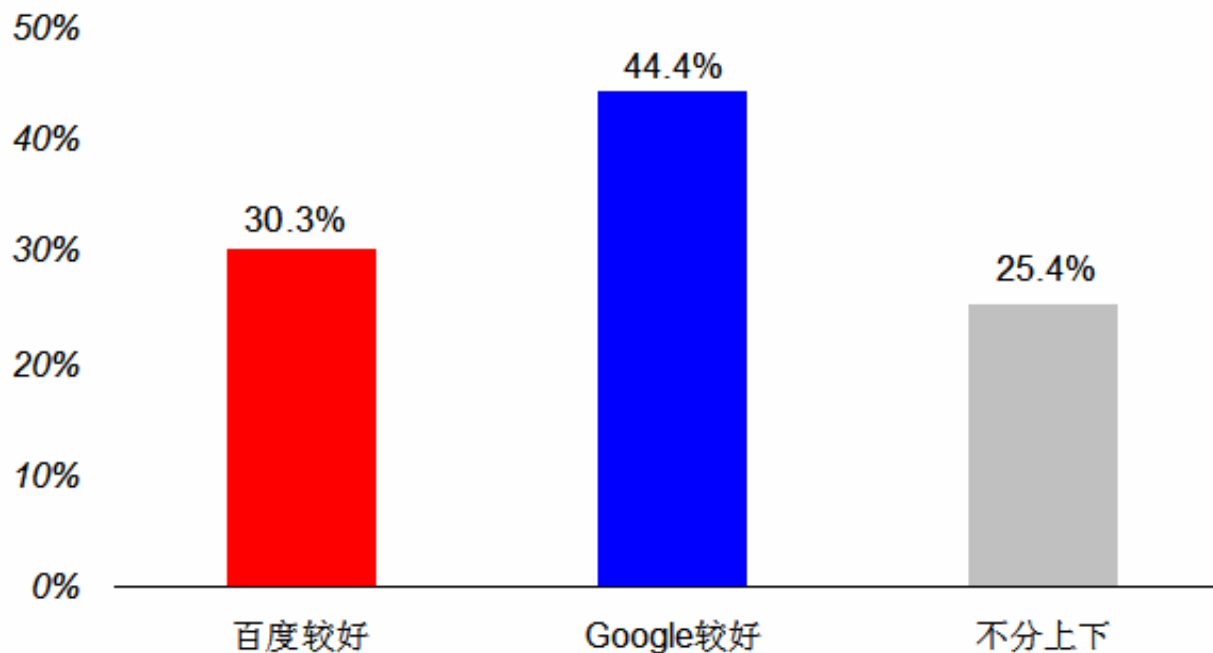


B3、B2、B1: 百度比Google好, 给百度搜索结果分别打3分、2分和1分;

G3、G2、G1: Google比百度好, 给Google搜索结果分别打3分、2分和1分;

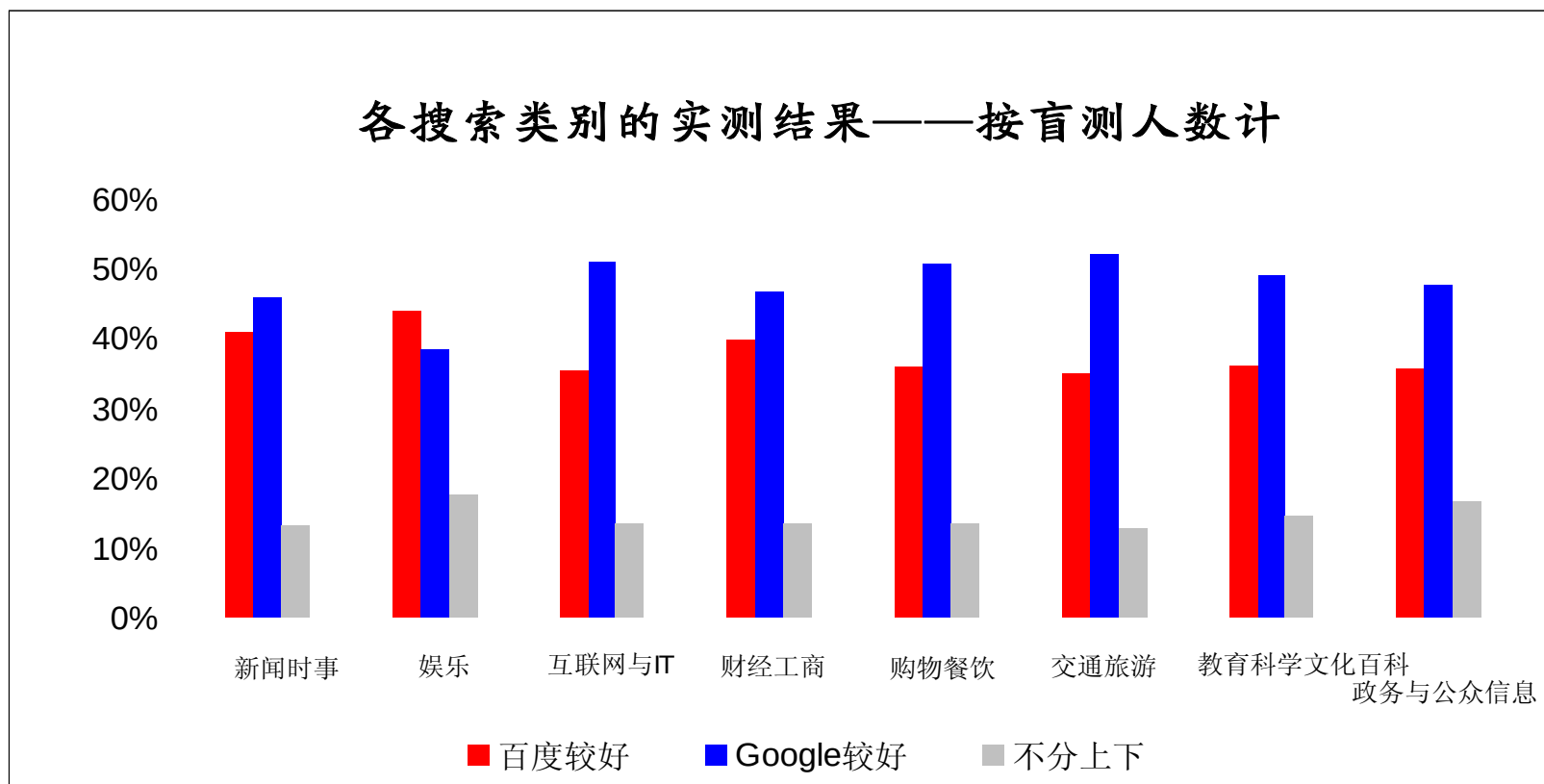
3分: 好很多; 1分: 只好一点

以盲测人数计算的盲测结果——总结果



每人最少4次的盲测结果中，如果百度较好的次数多于Google较好的次数，则将该盲测者计入“百度较好”，反之则计入“Google较好”。“一样好”包括百度与Google较好的次数正好一样多，以及盲测结果都是一样好这样两种情形

以盲测人数计算的盲测结果——分类结果



每人最少4次的盲测结果中，如果百度较好的次数多于Google较好的次数，则将该盲测者计入“百度较好”，反之则计入“Google较好”。“一样好”包括百度与Google较好的次数正好一样多，以及盲测结果都是一样好这样两种情形

Google胜出的幅度——按实测次数计算

（以下数据以Google胜出的百分点表示）

	学生	非学生	高端用户	总体
新闻时事	1.7	5.2	13.1	4.1
娱乐	-6.3	-4.3	0.5	-5.0
IT互联网	14.6	15.9	18.5	15.6
财经工商	2.3	9.8	18.9	8.2
购物、餐饮与生活服务	19.6	15.4	22.3	16.3
交通旅游	24.1	15.8	22.0	18.2
教育科学文化百科	13.3	12.1	14.9	12.5
政务与公众信息	-2.6	14.9	24.0	9.9
总体	6.8	9.1	15.6	8.4

Google胜出的幅度——按盲测人数计算

（以下数据以Google胜出的百分点表示）

	学生	非学生	高端用户	总体
新闻时事	3.0	5.6	9.4	4.8
娱乐	-7.2	-4.8	1.7	-5.5
IT互联网	15.6	15.6	18.3	15.6
财经工商	-2.7	9.6	15.4	6.9
购物、餐饮与生活服务	22.8	12.6	21.4	14.8
交通旅游	22.6	15.1	22.7	17.2
教育科学文化百科	13.0	13.2	17.0	13.1
政务与公众信息	1.5	16.0	26.1	12.0
总体	12.2	14.9	25.6	14.1

影响评测结果的用户特征因素

	百度较好	不分上下	Google较好	合计
性别:				
男	30.2%	25.4%	44.4%	100.0%
女	31.5%	24.5%	44.0%	100.0%
	30.3%	25.4%	44.4%	100.0%
年龄:				
18-24岁	 33.5%	24.6%	 41.9%	100.0%
25-30岁	29.5%	26.7%	43.8%	100.0%
31-40岁	24.8%	26.0%	 49.2%	100.0%
学历:				
高中	31.5%	28.4%	40.1%	100.0%
专科	30.0%	28.0%	42.1%	100.0%
本科	 31.2%	24.4%	44.4%	100.0%
硕士	26.5%	19.9%	53.6%	100.0%
博士	 15.9%	25.4%	 58.7%	100.0%

影响评测结果的用户上网行为因素

	百度较好	不分上下	Google较好	合计
上网频率：				
每周上网5天以上	 26.8%	25.2%	 45.1%	100.0%
每周上网不到5天	 34.3%	27.1%	 35.2%	100.0%
使用搜索频率：				
很频繁	 27.4%	24.5%	 48.1%	100.0%
比较频繁	 34.3%	24.1%	 41.6%	100.0%
一般	32.9%	31.2%	 35.9%	100.0%
偶尔用用	32.6%	34.9%	 32.6%	100.0%
搜索的重要性：				
10分	 26.3%	22.3%	 51.5%	100.0%
8到9分	 33.0%	23.6%	 43.4%	100.0%
1到7分	31.1%	27.3%	 41.6%	100.0%

搜索质量盲测评估的判断依据（1）

（占全部实测次数的比例，多选结果）

	百度好	Google好	总体
相关性更好	53.7%	59.6%	57.0%
资料来源丰富	34.8%	38.5%	36.8%
链接摘要描述精准	25.1%	26.8%	26.0%
资料更新及时	26.1%	24.7%	25.3%
中文分词正确性高	18.9%	17.8%	18.3%
结果中垃圾网页少	11.0%	14.5%	13.0%
结果中重复内容少	10.4%	9.5%	9.9%
结果中作弊网页少	7.0%	9.4%	8.3%
结果中死链少	4.7%	4.8%	4.8%
其他	5.6%	6.2%	5.9%
我是凭感觉作出的判断	25.2%	21.7%	23.3%
判断依据的平均引用个数	2.23	2.33	2.28

搜索质量盲测评估的判断依据（2）

（占全部实测次数的比例，多选结果）

	B3	B2	B1	G1	G2	G3
相关性更好	60.5%	51.9%	40.1%	41.0%	61.3%	65.7%
资料来源丰富	39.4%	33.1%	26.3%	27.0%	37.4%	43.3%
链接摘要描述精准	29.4%	24.2%	16.2%	13.3%	24.2%	33.2%
资料更新及时	31.8%	23.8%	15.6%	13.9%	20.0%	31.4%
中文分词正确性高	22.1%	18.5%	11.8%	9.6%	16.0%	21.8%
结果中垃圾网页少	13.1%	11.0%	6.0%	8.7%	11.5%	18.4%
结果中重复内容少	12.0%	9.9%	7.4%	6.4%	7.6%	11.7%
结果中作弊网页少	9.4%	6.1%	2.8%	5.1%	6.2%	12.7%
结果中死链少	6.5%	3.2%	2.9%	1.9%	3.0%	6.9%
其他	5.9%	5.6%	5.1%	6.3%	6.2%	6.1%
我是凭感觉作出的判断	23.7%	23.8%	30.6%	30.3%	20.8%	19.0%
判断依据的平均引用个数	2.54	2.11	1.65	1.63	2.14	2.70



盲测者使用搜索引擎的 习惯与态度

使用的搜索引擎不满意的地方

对百度不满意的地方（前四位）	
百度将广告与自然搜索结果混排在一起	35.7%
百度不完全提供来自中国境外的中文搜索结果	27.3%
百度搜索结果不够客观公正	25.4%
百度搜索结果的相关性不好	23.5%
对Google不满意的地方（前四位）	
Google搜索不稳定	30.3%
Google网页快照不能用	29.1%
Google不能提供MP3搜索	27.7%
Google不能提供（百度）贴吧	23.5%



有关正望咨询的最新调查报告 请联系：

北京正望咨询有限公司

China IntelliConsulting Corporation

sales@iaskchina.com

(010)8256-2129