



(12) 发明专利申请

(10) 申请公布号 CN 103064945 A

(43) 申请公布日 2013. 04. 24

(21) 申请号 201210575284. 8

(22) 申请日 2012. 12. 26

(71) 申请人 吉林大学

地址 130012 吉林省长春市前进大街 2699  
号

(72) 发明人 左万利 赫枫龄 王俊华 王鑫  
凤丽洲 王英 彭涛 万海旭  
苏雪阳 高宁宁 闫昭 张雪松

(74) 专利代理机构 长春吉大专利代理有限责任  
公司 22201

代理人 齐安全 胡景阳

(51) Int. Cl.

G06F 17/30 (2006. 01)

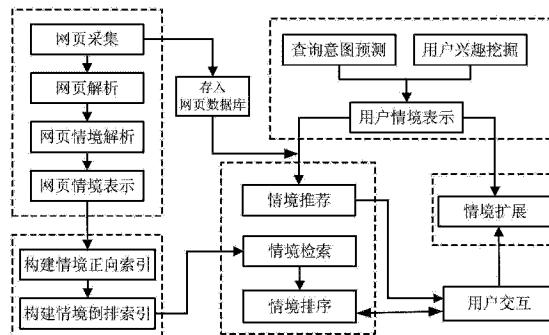
权利要求书5页 说明书23页 附图9页

(54) 发明名称

基于本体的情境搜索方法

(57) 摘要

本发明公开了一种基于本体的情境搜索方法，旨在克服现有技术存在精准度不够、个性化不足、智能性缺失与情境被忽视等问题。该方法的步骤为：1. 由网页情境解析与表示模块对采集到的网页进行解析并表示网页情境信息；2. 由用户情境挖掘与表示模块实现用户情境挖掘与表示，并和第1步骤同时进行；3. 由情境索引模块设计融入网页情境信息的索引结构；4. 由情境扩展模块实现情境扩展即形成查询情境，所述的情境扩展是指利用用户情境，在情境级别上对查询词进行扩展，限定查询词所属的语境，形成查询情境；5. 由情境查询处理模块实现情境查询处理：其步骤为：1) 情境推荐；2) 情境检索；3) 情境排序；排序之后的结果就可以提供到用户查询界面中。



1. 一种基于本体的情境搜索方法,其特征在于,所述的基于本体的情境搜索方法包括如下步骤:

1) 由网页情境解析与表示模块对采集到的网页进行解析并表示网页情境信息:

(1) 使用网络爬虫即 Crawler 采集网页并存入网页数据库;

(2) 建立 DOM 树进行网页解析,并使用实时搜索爬虫提取网页最近更新时间;

(3) 实现网页情境解析;

(4) 利用命名实体识别技术获取文档内的时间与地点信息,并将文档表示为网页情境形式:

$PC = \langle \{W, C\}, PG, PT \rangle$

其中 :W 为词或术语, C 为语境词, PG 为网页内的地点信息, PT 为网页时间信息, 包含创建时间、更新时间与网页内时间;

2) 由用户情境挖掘与表示模块实现用户情境挖掘与表示并和第 1) 步骤同时进行:

(1) 用户兴趣挖掘;

(2) 查询意图预测;

(3) 基于步骤(1)与步骤(2)中得到的处理结果, 使用本体技术给出清晰、明确、可计算的用户情境说明和描述; 用户情境包括用户兴趣、查询意图以及搜索时间和地点; 可以将用户情境表示为:

$UC = \langle L, S, I, Q, UG, UT \rangle$

其中 :L 为用户长期兴趣, S 为用户短期兴趣, I 为当前查询意图, Q 为检索词, UG 为用户地理位置, UT 为用户查询时间;

3) 由情境索引处理模块设计融入网页情境信息的索引结构:

(1) 构建情境正向索引;

(2) 构建情境倒排索引;

4) 由情境扩展模块实现情境扩展即形成查询情境;

5) 由情境查询处理模块实现情境查询处理。

2. 按照权利要求 1 所述的基于本体的情境搜索方法,其特征在于,所述的实现网页情境解析包括如下步骤:

1) 如果网页是中文网页,则对网页进行分词;

2) 利用通用本体和自然语言处理技术构建语义关联图,对词或术语进行词义标注;

3) 确定语义关联图中语义关系权重;语义关系权重的确定采用如下公式:

$$SR(Sen_1, Sen_2) = \max_r - \frac{(\max_r - \min_r)}{n_r(Sen_1)}$$

其中 : $\max_r$ 、 $\min_r$  是赋予关系 r 的最大权重与最小权重,  $n_r(Sen_1)$  是在语义关联图中从词义  $Sen_1$  出发的关系 r 的边的个数;

4) 计算歧义词 W 的每个义项和上下文词集的关联度,公式为:

$$\sum_{j=1}^{|CS|} WCR(WS_i, CS)$$

其中 : $WS_i$  为歧义词 W 的第 i 个义项, CS 为满足以下条件的义项集合:①出现在语义关

联图中,②为上下文词集中词的一个义项,WCR(WS<sub>i</sub>, CS<sub>j</sub>) 为基于语义关联图和语义关系权重计算义项 WS<sub>i</sub> 和 CS<sub>j</sub> 的关联度函数;

5)选取关联度最大的义项作为词 W 在上下文中的释义,并提取释义的直接上位概念作为语境词,获取 W 的语境,从而实现文档的情境解析。

3. 按照权利要求 1 所述的基于本体的情境搜索方法,其特征在于,所述的用户兴趣挖掘包括如下步骤:

1) 对用户搜索日志进行挖掘,获取用户点击的网页类别;

2) 量化用户点击各类网页的会话数和频度;

3) 设定阈值,依据频度,提取用户长期兴趣概念集 C<sub>L</sub> 和短期兴趣概念集 C<sub>S</sub>;

用户长期兴趣与短期兴趣分别用通用本体的局部来表示,并可动态漂移;针对用户兴趣的漂移性,用户的长期兴趣和短期兴趣漂移表示为:

$L' = \text{long\_term\_interest\_drift}(L, S, H)$

$S' = \text{short\_term\_interest\_drift}(S, L, H)$

其中:L 为用户长期兴趣, S 为用户短期兴趣, H 为查询历史数据, L' 和 S' 分别为用户兴趣增量学习过程获得的长期兴趣和短期兴趣;

4) 以用户兴趣为基础,基于通用本体,提取 C<sub>L</sub> 和 C<sub>S</sub> 中的上下位词汇链,将各词汇链衔接在一起,即得用户兴趣模型。

4. 按照权利要求 1 所述的基于本体的情境搜索方法,其特征在于,所述的查询意图预测包括如下步骤:

1) 根据用户本次的检索词,从用户连续的搜索行为中预测查询意图,并实现用户情境获取确定用户的查询意图;采用动态贝叶斯网络推断当前用户查询的搜索意图,形式化表示为:

$I = \text{Intent\_infer}(L, S, P, Q) = \langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$

其中:I 为用户的查询意图, p<sub>1</sub>, p<sub>2</sub>, ..., p<sub>n</sub> 分别为查询子意图 I<sub>1</sub>, I<sub>2</sub>, ..., I<sub>n</sub> 的概率,L 为用户的长期兴趣,S 为用户的短期兴趣,P 为点击网页,Q 为查询,Intent\_infer 为基于动态贝叶斯网的查询意图推测函数;

2) 通过动态贝叶斯网络,利用搜索系统的先验知识,计算用户当前查询意图的后验概率分布;根据已有的经验知识,定义了①用户的查询意图 I、②用户的长期兴趣 L、③用户的短期兴趣 S、④点击网页 P 与⑤查询 Q 五个数据变量,并找出了数据变量之间的因果关系,获得动态贝叶斯网络;

查询意图预测涉及的主要概率计算公式如下:

$$p(I_t | I_{t-1}, L_t, S_t, P_{t-1}, Q_t) = \frac{p(I_t, I_{t-1}, L_t, S_t, P_{t-1}, Q_t)}{p(I_{t-1}, L_t, S_t, P_{t-1}, Q_t)} = \frac{p(I_t)p(I_{t-1}|I_t)p(L_t|I_t)p(S_t|I_t)p(P_{t-1}|I_t)p(Q_t|I_t)}{p(I_{t-1})p(L_t)p(S_t)p(P_{t-1})p(Q_t)}$$

其中:I<sub>t</sub> 表示第 t 次查询的意图, I<sub>t-1</sub> 表示第 t-1 次查询的意图, L<sub>t</sub> 表示第 t 次查询时的用户长期兴趣, S<sub>t</sub> 表示第 t 次查询时的用户短期兴趣, P<sub>t-1</sub> 表示第 t-1 次查询时的点击的网页, Q<sub>t</sub> 表示第 t 次查询。

5. 按照权利要求 1 所述的基于本体的情境搜索方法,其特征在于,所述的构建情境正向索引是指:在传统正向索引的基础上,设计融入了语境词和时间、空间信息的情境正向索引结构;

所述的构建情境倒排索引是指：利用情境正向索引，设计了〈词，语境词〉到网页列表的语境索引结构，并借助时空数据库技术，建立基于 R\* 树的空间索引和基于 B+ 树的时间索引，最后将语境索引与时空索引有机地结合在一起，得到真正直接支持情境检索的索引结构。

6. 按照权利要求 1 所述的基于本体的情境搜索方法，其特征在于，所述的情境扩展是指利用用户情境，在情境级别上对查询词进行扩展，限定查询词所属的语境，形成查询情境；

利用通用本体结合查询词和权利要求 1 中步骤 2) 中得到的查询意图获得相应语境词，实现情境扩展，形式化表示为：

$\text{SearchContext} = \text{ContextExpansion}(Q, O, I, \text{UG}, \text{UT})$

其中 : $Q$  为查询词， $O$  为通用本体， $I$  为查询意图， $\text{UG}$  为查询地点， $\text{UT}$  为查询时间，在绝大多数情况下，本发明推荐的语境词是恰当的，但当用户意图发生突变时，最佳匹配的语境词不一定是本次搜索所期望的，此时本发明将提供次佳匹配的语境词；因此，需要将多个语境词按匹配情况进行排序；

1) 根据用户提供的检索词和在权利要求 1 中步骤 2) 中用户情境挖掘与表示模块所获得的用户情境信息，使用通用本体提取查询词  $Q$  的若干释义  $\langle Q_S_1, Q_S_2, \dots, Q_S_m \rangle$  及相应的若干语境  $\langle C_1, C_2, \dots, C_m \rangle$ ；

2) 在权利要求 1 中步骤 2) 中用户情境挖掘与表示模块所获得的查询意图  $\langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$  的基础上，利用朴素贝叶斯模型计算语境词  $C_i$  的后验概率  $p(C_i | I_1, I_2, \dots, I_n)$ ，将多个语境词按匹配情况进行降序排序，并按照该值的降序向用户推荐相应语境的情境扩展，其中最大值所对应的语境词为最佳语境词，其次为次佳语境词；

3) 若用户选择某情境扩展，则取相应语境为最终确认结果，否则自动选取  $p(C_i | I_1, I_2, \dots, I_n)$  值最大的语境为最终确认结果，且以次佳匹配的语境词为备选，用户选定扩展的语境词后，如果用户选定的语境词与本发明提供的最佳语境词不一致，需要更新该用户的兴趣和查询意图；

4) 最后获取查询词  $Q$  的同义词集  $\text{Syn}$ ，查询情境形式化表示为： $\langle \{ \langle Q, \text{Syn}, C \rangle \}, \text{UG}, \text{UT} \rangle$ ，其中 : $Q$  为查询词， $\text{Syn}$  为  $Q$  的同义词集， $C$  为  $Q$  的语境词， $\text{UG}$  为查询地点， $\text{UT}$  为查询时间。

7. 按照权利要求 1 所述的基于本体的情境搜索方法，其特征在于，所述的情境查询处理模块实现情境查询处理，步骤如下：

1) 情境推荐：

在用户打开查询界面但是没有进行查询操作时，对于经权利要求 1 中步骤 1) 网页采集存储在网页数据库中网页，根据网页情境与用户情境的匹配程度及网页权威性，以合理的顺序向用户主动推荐网页，达到“不搜亦得”的智能化检索目标；情境推荐的形式化表示为：

$\text{Result} = \text{Recommend}(\text{Pages}, \text{UT}, \text{UG}, A, R_{\text{SL}})$

其中 : $\text{Pages}$  表示新抓取的网页集合， $\text{UT}$  表示用户最近查询时间， $\text{UG}$  表示用户所处地点， $A$  表示网页权威性， $R_{\text{SL}}$  表示用户兴趣与网页情境的相关性；

情境推荐形式化表示中的各项的计算方法如下：

a. 网页权威性  $A$  的计算采用经典的基于网页链接的 PageRank 静态排序算法；

b. 用户兴趣与网页情境的相关性  $R_{SL} = m \times R_s + n \times R_L$ ; 其中 :m 和 n 是权衡系数,  $R_s$  表示短期兴趣与网页情境的相关性,  $R_L$  表示长期兴趣与网页情境的相关性;

$$R_s = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Sset|} p(Cset_i | Sset_j)}{|Cset| \cdot |Sset|} \quad R_L = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Lset|} p(Cset_i | Lset_j)}{|Cset| \cdot |Lset|}$$

其中 : $Sset$  表示短期兴趣概念集合,  $Lset$  表示长期兴趣概念集合,  $Cset$  表示网页情境中的语境词集;

## 2) 情境检索 :

此过程发生在用户提交了此次查询的查询词之后,

(1) 经权利要求 1 中步骤 4) 情境扩展模块处理获得语境词、地理位置、时间信息, 利用这些信息, 在权利要求 1 的步骤 3) 中构建的情境索引数据库中进行检索;

(2) 同时访问权利要求 1 中步骤 1) 建立的网页数据库, 得到相匹配的网页, 形式化表示为:

Result=Search(Index, Q, C, Syn, UG, UT)

用 C 和 Q 或 C 和 Syn 在索引 Index 中的第一层语境索引 CI 中进行检索, 用用户所处地点 UG 在第二层空间索引 GI 中进行检索, 用用户最近检索时间 UT 在第三层索引时间 TI 中进行检索; 在每一层检索的具体过程描述如下:

### a. 在语境层检索 :

构建查询词的语境检索项, 语境检索既利用消解歧义后的语境词实现精准检索, 又通过查询词的同义词集实现语义相关检索, 同义词与查询词具有共同的语境词, 这样每个查询词  $Q_i$  产生的语境检索项表示为:  $\{ \langle Q_i, C_i \rangle, \langle Syn_{i1}, C_i \rangle, \dots, \langle Syn_{ik}, C_i \rangle \}$ ,

其次, 在语境索引中, 通过查询词的语境检索项检索相应的语境索引项, 进而在空间索引中继续检索;

### b. 在空间层检索 :

借助地理学本体和地理信息系统将查询中涉及空间信息的查询词项映射为区域或坐标, 并采用自然语言处理技术识别出与空间信息相关的语义信息, 从而实现查询词空间敏感信息的检测;

根据检测结果进一步判断, 若查询词不具有空间敏感性, 检索的内容包括语境索引项下的所有空间索引项; 否则, 利用用户所在位置的空间信息 UG 和查询词项涉及的空间敏感信息, 确定适合空间检索的窗口大小, 从而在空间索引 GI 中检索符合用户空间信息需求的相关索引项, 进而在时间索引中继续检索;

### c. 在时间层检索 :

对于时间不敏感查询, 其检索范围为整个时间轴; 对于时间敏感查询, 检索范围为  $[st, et]$  表达的时间区间, st 为起始时间, et 为终止时间, 是经过情境扩展模块处理用户提交查询词等信息分析后得到的, 例如, 最近 1 小时或 1 天或 1 周或 1 月或 1 年可以统一表示为过去起始时间到当前时间的时间区间, 用 UT 在 B+ 树 TI 中进行检索, 将落在 UT 范围内的网页作为返回结果;

## 3) 情境排序 :

在经情境检索取得返回结果之后,对于情境检索返回的结果,依据情境相关性及网页权威性进行排序,排序之后的结果就可以提供到用户查询界面中了,情境排序的形式化表示为:

$$\text{Result}' = \text{Rank}(\text{Result}, A, R)$$

其中:A 为网页权威性,采用经典的基于网页链接的 PageRank 静态排序算法, R 为情境相关性,是用户情境表示  $UC = \langle L, S, I, Q, UG, UT \rangle$  与网页情境表示  $PC = \langle \{W, C\}, PG, PT \rangle$  匹配的结果,即:  $R = \text{Match}(UC, PC)$ , 该公式基于 Q 与 W 的匹配、I 与 C 的匹配、UG 与 PG 的匹配、UT 与 PT 匹配,返回相应网页的情境相关性值;

本发明通过实验确定参数 A 和 R 在最终排序算法中的权重,并依据该权重对经典排序结果进行调整;

(1) 采用经典的基于网页链接的 PageRank 静态排序算法,确定网页权威性 A 的值;

(2) 使用 BM25 概率检索模型,通过计算网页情境中的语境词与查询意图间的相关性概率,获得网页面档与查询词之间的概念相关性得分  $Rc$ ,即在查询词扩展所得的各上位概念,与网页情境解析所得的各上位概念之间,计算相关性概率,文档的分值可以匹配到各概念项概率求和;

(3) 网页情境解析的某上位概念  $C_x$  的概率,可以由相关集合中所有扩展出  $C_x$  的词的频率相加求和得到;

(4) 通过比对  $UG$  与  $PG$  中各地理位置信息,获得  $UG$  和  $PG$  的匹配度得分  $Rg$ ,计算与查询词匹配之后所得向量的分量进行加权求和得出,即  $\sum g_i \times w_i$ ,  $g_i$  表示地理位置信息的向量各分量,  $w_i$  为对应的权重,取值区间为  $[0, 1]$ ;

(5) 索引项的时间信息得分  $Rt$ ,由网页情境时间信息是否匹配查询意图要求的信息,如果匹配取值为 1,否则取值为 0,乘以权值常数 0.85 以保证其影响力,再加上由查询时间和网页发布时间差值函数  $f(t)$  确定的网页新鲜程度得分,其中  $f(t) = \text{DecayRate}^{(St-Pt)}$ ,  $St$  为查询时间,  $Pt$  为网页发布时间,  $\text{DecayRate}$  取值为 0.5;

(6) 最终由  $Rc+Rg+Rt$  得出 R 的值;

(7) 根据计算结果进行排序,排序之后的结果就可以提供到用户查询界面中。

## 基于本体的情境搜索方法

### 技术领域

[0001] 本发明涉及一种搜索引擎领域的搜索方法,更确切地说,本发明涉及一种基于通用本体的情境搜索方法。

### 背景技术

[0002] 现如今,海量的信息资源以网络为载体,人们已习惯于通过搜索获得信息。搜索改变了我们寻找答案、文章、广告、产品、人物和地点等诸多事物的方式,影响着我们的学习、工作和生活,为我们的决策提供参考信息。然而人们在获得丰富信息同时,也面临着信息选择方面的困惑,且通常需要反复修改查询词来重新定位自己所需要的信息。

[0003] 搜索引擎是指根据一定的策略、运用特定的计算机程序从特定网络如互联网上搜集信息,在对信息进行组织和处理后,为用户提供检索服务,将与用户检索相关的信息展示给用户的系统。

[0004] 现在搜索引擎所面临的问题并不是信息量不足,而是信息资源的泛滥让人们无法快捷、精确的找到真正符合自己需要的信息,大量用户认为传统的搜索引擎返回的结果中垃圾信息太多,无法提供优质的搜索服务,其不足主要体现在:

[0005] 1. 精准度不够

[0006] 据权威统计,40%的搜索结果不能回答用户的问题,50%的查询需在搜索过程中不断调整查询词以试探最佳结果。

[0007] 2. 个性化不足

[0008] 传统搜索引擎没有综合考虑用户的搜索历史、当前搜索内容、兴趣、所处位置和时间,千人一面,无法提供个性化的搜索服务。

[0009] 3. 智能性缺失

[0010] 传统搜索引擎很少动态追踪用户行为、挖掘检索词背后隐含兴趣和意图,无法智能化完善用户查询以辅助用户完成搜索任务,不能向用户自动推荐其所需信息。

[0011] 4. 语境被忽视

[0012] 传统搜索引擎忽略了词汇的语境扩展,只能根据查询词与网页词汇的匹配进行简单地匹配和检索。

[0013] 传统搜索引擎将网页简单表示为若干词汇的向量,这种表示缺少网页情境信息(如:词、语境词、网页地点、更新时间等)。另外,传统搜索引擎在搜索过程中没有考虑用户情境(如:与搜索行为有关的时间、地点、用户兴趣和查询意图),目前获取用户情境的主要方法有两种:

[0014] 一是由用户明确提供,这种方法直接有效,但会增加用户负担,研究表明,用户并不愿意提供这种反馈信息;

[0015] 二是通过系统对交互数据进行计算推理获得,这种方法符合搜索智能化发展战略目标,但如何自动获取准确有效的用户情境也是一个不可回避的问题。另一方面当前搜索系统多数采用词到文档的倒排索引,这种索引结构简单、操作方便、检索快速、吞吐开销较

小,但缺少网页情境信息,且不利于用户情境的应用,在传统索引基础上进行检索只能遵循“先搜索、再过滤”的技术路线,效率低,效果差,且传统索引基础上进行检索的过程中大部分用户情境信息都不能直接影响检索结果,通常用户提交的搜索词是多义的,例如:对于搜索词 apple,可能有水果、电子产品、电影等多种释义,在该情况下,传统检索模型会返回很多冗余记录。在传统结果排序方法中,比较流行的是词频统计法和超链接分析法。这些方法忽略了搜索个体之间搜索情境的差异,导致当不同用户提交相同查询时,返回相同的搜索结果,且不能依据用户的后续行为,重新调整搜索结果。

[0016] 由此可见传统搜索引擎脱离情境是以上诸多问题产生的根源。因此,如何将情境融入搜索是一个亟待解决的问题。当前,较完整的情境搜索定义是:情境搜索包含 7 个要素(6W&1H),它强调“以人(Who)为本”,也就是以用户为中心,根据其搜索行为的时间(When)、地点(Where)、输入(What)、需求(Want)、习惯(How)、背景(Why)等因素,得到最适合的搜索结果,再将这一结果直接呈现给用户。情境具有多维性、多样性、动态性和延展性等特征:

[0017] 1. 多维性

[0018] 用户情境包括与搜索行为有关的时间、地点、兴趣、意图等,网页情境包括词、语境词、网页地点、更新时间等,这些信息可以全方位多角度地描述情境,体现了情境的多维性;

[0019] 2. 多样性

[0020] 情境搜索的服务对象千差万别,情境搜索要针对不同用户提供个性化的检索服务;

[0021] 3. 动态性

[0022] 用户情境中兴趣、意图的漂移,网页情境中信息、热点的更新,体现出了情境明显的动态性;

[0023] 4. 延展性

[0024] 在用户情境中,当前查询意图通常与其前数次查询意图相关,即用户查询意图具有短期延展性。

[0025] 目前信息表示的主要方法有词袋、向量和图,但均不足以应对用户情境表示面临的各种挑战,而本体是关于领域知识的概念化、形式化的明确规范,定义了组成主题领域的词汇表的基本术语及其关系,以及结合这些术语和关系来定义词汇表外延的规则。经过仔細分析,本方法采用的本体技术作为一种通用工具,特别适合情境搜索问题。

## 发明内容

[0026] 本发明所要解决的技术问题是克服了现有技术存在精准度不够、个性化不足、智能性缺失与情境被忽视等问题,提供了一种基于本体的情境搜索方法。

[0027] 为解决上述技术问题,本发明是采用如下技术方案实现的:所述的基于本体的情境搜索方法包括如下步骤:

[0028] 1. 由网页情境解析与表示模块对采集到的网页进行解析并表示网页情境信息;

[0029] (1) 使用网络爬虫即 Crawler 采集网页并存入网页数据库;

[0030] (2) 建立 DOM 树进行网页解析,并使用实时搜索爬虫提取网页最近更新时间;

[0031] (3) 实现网页情境解析；  
[0032] (4) 利用命名实体识别技术获取文档内的时间与地点信息，并将文档表示为网页情境形式：

[0033]  $PC = \langle \{W, C\}, PG, PT \rangle$   
[0034] 其中 :W 为词或术语, C 为语境词, PG 为网页内的地点信息, PT 为网页时间信息, 包含创建时间、更新时间与网页内时间。  
[0035] 2. 由用户情境挖掘与表示模块实现用户情境挖掘与表示并和第 1 步骤同时进行：  
[0036] (1) 用户兴趣挖掘；  
[0037] (2) 查询意图预测；  
[0038] (3) 基于步骤 (1) 与步骤 (2) 中的得到的处理结果, 使用本体技术给出清晰、明确、可计算的用户情境说明和描述 ; 用户情境包括用户兴趣、查询意图以及搜索时间和地点 ; 可以将用户情境表示为 :

[0039]  $UC = \langle L, S, I, Q, UG, UT \rangle$   
[0040] 其中 :L 为用户长期兴趣, S 为用户短期兴趣, I 为当前查询意图, Q 为查询词, UG 为用户地理位置, UT 为用户查询时间；

[0041] 3. 由情境索引处理模块设计融入网页情境信息的索引结构：  
[0042] (1) 构建情境正向索引；  
[0043] (2) 构建情境倒排索引；  
[0044] 4. 由情境扩展模块实现情境扩展并形成查询情境；  
[0045] 5. 由情境查询处理模块实现情境查询处理。

[0046] 技术方案中所述的实现网页情境解析包括如下步骤：  
[0047] 1) 如果网页是中文网页, 则对网页进行分词；  
[0048] 2) 利用通用本体和自然语言处理技术构建语义关联图, 对词或术语进行词义标注；  
[0049] 3) 确定语义关联图中语义关系权重 ; 语义关系权重的确定采用如下公式：

$$SR(Sen_1, Sen_2) = \max_r - \frac{(\max_r - \min_r)}{n_r(Sen_1)}$$

[0050] 其中 : $\max_r$ 、 $\min_r$  是赋予关系 r 的最大权重与最小权重,  $n_r(Sen_1)$  是在语义关联图中从词义  $Sen_1$  出发的关系 r 的边的个数；

[0051] 4) 计算歧义词 W 的每个义项和上下文词集的关联度, 公式为：  

$$\sum_{j=1}^{|CS|} WCR(WS_i, CS_j)$$

[0052] 其中 : $WS_i$  为歧义词 W 的第 i 个义项, CS 为满足以下条件的义项集合 :①出现在语义关联图中, ②为上下文词集中词的一个义项,  $WCR(WS_i, CS_j)$  为基于语义关联图和语义关系权重计算义项  $WS_i$  和  $CS_j$  的关联度函数；

[0053] 5) 选取关联度最大的义项作为词 W 在上下文中的释义, 并提取释义的直接上位概念作为语境词, 获取 W 的语境, 从而实现文档的情境解析。

[0054] 技术方案中所述的用户兴趣挖掘包括如下步骤：

[0055] 1) 对用户搜索日志进行挖掘, 获取用户点击的网页类别；

[0058] 2) 量化用户点击各类网页的会话数和频度；

[0059] 3) 设定阈值，依据频度，提取用户长期兴趣概念集  $C_L$  和短期兴趣概念集  $C_S$ ；

[0060] 用户长期兴趣与短期兴趣分别用通用本体的局部来表示，并可动态漂移；针对用户兴趣的漂移性，用户的长期兴趣和短期兴趣漂移表示为：

[0061]  $L' = \text{long\_term\_interest\_drift}(L, S, H)$

[0062]  $S' = \text{short\_term\_interest\_drift}(S, L, H)$

[0063] 其中： $L$  为用户长期兴趣， $S$  为用户短期兴趣， $H$  为查询历史数据， $L'$  和  $S'$  分别为用户兴趣增量学习过程获得的长期兴趣和短期兴趣；

[0064] 4) 以用户兴趣为基础，基于通用本体，提取  $C_L$  和  $C_S$  中的上下位词汇链，将各词汇链衔接在一起，即得用户兴趣模型。

[0065] 技术方案中所述的查询意图预测包括如下步骤：

[0066] 1) 根据用户本次的查询词，从用户连续的搜索行为中预测查询意图，并实现用户情境获取确定用户的查询意图；采用动态贝叶斯网络推断当前用户查询的搜索意图，形式化表示为：

[0067]  $I = \text{Intent\_infer}(L, S, P, Q) = \langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$

[0068] 其中： $I$  为用户的查询意图， $p_1, p_2, \dots, p_n$  分别为查询子意图  $I_1, I_2, \dots, I_n$  的概率， $L$  为用户的长期兴趣， $S$  为用户的短期兴趣， $P$  为点击网页， $Q$  为查询， $\text{Intent\_infer}$  为基于动态贝叶斯网的查询意图推测函数；

[0069] 2) 通过动态贝叶斯网络，利用搜索系统的先验知识，计算用户当前查询意图的后验概率分布；根据已有的经验知识，定义了①用户的查询意图  $I$ 、②用户的长期兴趣  $L$ 、③用户的短期兴趣  $S$ 、④点击网页  $P$  与⑤查询  $Q$  五个数据变量，并找出了数据变量之间的因果关系，获得动态贝叶斯网络；

[0070] 查询意图预测涉及的主要概率计算公式如下：

[0071]

$$p(I_t | I_{t-1}, L_t, S_t, P_{t-1}, Q_t) = \frac{p(I_t, I_{t-1}, L_t, S_t, P_{t-1}, Q_t)}{p(I_{t-1}, L_t, S_t, P_{t-1}, Q_t)} = \frac{p(I_t)p(I_{t-1}|I_t)p(L_t|I_t)p(S_t|I_t)p(P_{t-1}|I_t)p(Q_t|I_t)}{p(I_{t-1})p(L_t)p(S_t)p(P_{t-1})p(Q_t)}$$

[0072] 其中： $I_t$  表示第  $t$  次查询的意图， $I_{t-1}$  表示第  $t-1$  次查询的意图， $L_t$  表示第  $t$  次查询时的用户长期兴趣， $S_t$  表示第  $t$  次查询时的用户短期兴趣， $P_{t-1}$  表示第  $t-1$  次查询时的点击的网页， $Q_t$  表示第  $t$  次查询。

[0073] 技术方案中所述的构建情境正向索引是指：在传统正向索引的基础上，设计融入了语境词和时间、空间信息的情境正向索引结构；

[0074] 所述的构建情境倒排索引是指：利用情境正向索引，设计了〈词，语境词〉到网页列表的语境索引结构，并借助时空数据库技术，建立基于 R\* 树的空间索引和基于 B+ 树的时间索引，最后将语境索引与时空索引有机地结合在一起，得到真正直接支持情境检索的索引结构。

[0075] 技术方案中所述的情境扩展是指利用用户情境，在情境级别上对查询词进行扩展，限定查询词所属的语境，形成查询情境；

[0076] 利用通用本体结合查询词和权利要求 1 中步骤 2) 中得到的查询意图获得相应语境词，实现情境扩展，形式化表示为：

[0077] SearchContext = ContextExpansion(Q, 0, I, UG, UT)

[0078] 其中 :Q 为查询词, 0 为通用本体, I 为查询意图, UG 为查询地点, UT 为查询时间。在绝大多数情况下, 本发明推荐的语境词是恰当的, 但当用户意图发生突变时, 最佳匹配的语境词不一定是本次搜索所期望的, 此时本发明将提供次佳匹配的语境词; 因此, 需要将多个语境词按匹配情况进行排序;

[0079] 1) 根据用户提供的查询词和在本方法步骤 2 中用户情境挖掘与表示模块所获得的用户情境信息, 使用通用本体提取查询词 Q 的若干释义  $\langle QS_1, QS_2, \dots, QS_m \rangle$  及相应的若干语境  $\langle C_1, C_2, \dots, C_m \rangle$ ;

[0080] 2) 在本方法步骤 2 中用户情境挖掘与表示模块所获得的查询意图  $\langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$  的基础上, 利用朴素贝叶斯模型计算语境词  $C_i$  的后验概率  $p(C_i | I_1, I_2, \dots, I_n)$ , 将多个语境词按匹配情况进行降序排序, 并按照该值的降序向用户推荐相应语境的情境扩展, 其中最大值所对应的语境词为最佳语境词, 其次为次佳语境词。

[0081] 3) 若用户选择某情境扩展, 则取相应语境为最终确认结果, 否则自动选取  $p(C_i | I_1, I_2, \dots, I_n)$  值最大的语境为最终确认结果, 且以次佳匹配的语境词为备选。用户选定扩展的语境词后, 如果用户选定的语境词与本发明提供的最佳语境词不一致, 需要更新该用户的兴趣和查询意图;

[0082] 4) 最后获取查询词 Q 的同义词集 Syn, 查询情境形式化表示为:  $\langle \{ \langle Q, Syn, C \rangle \}, UG, UT \rangle$ , 其中 :Q 为查询词, Syn 为 Q 的同义词集, C 为 Q 的语境词, UG 为查询地点, UT 为查询时间。

[0083] 技术方案中所述的情境查询处理模块实现情境查询处理, 步骤如下:

[0084] 1) 情境推荐:

[0085] 在用户打开查询界面但是没有进行查询操作时, 对于经权利要求 1 中步骤 1) 网页采集存储在网页数据库中网页, 根据网页情境与用户情境的匹配程度及网页权威性, 以合理的顺序向用户主动推荐网页, 达到“不搜亦得”的智能化检索目标; 情境推荐的形式化表示为:

[0086] Result = Recommend(Pages, UT, UG, A, R<sub>SL</sub>)

[0087] 其中 :Pages 表示新抓取的网页集合, UT 表示用户最近查询时间, UG 表示用户所处地点, A 表示网页权威性, R<sub>SL</sub> 表示用户兴趣与网页情境的相关性。

[0088] 情境推荐形式化表示中的各项的计算方法如下:

[0089] a. 网页权威性 A 的计算采用经典的基于网页链接的 PageRank 静态排序算法;

[0090] b. 用户兴趣与网页情境的相关性  $R_{SL} = m \times R_s + n \times R_l$ ; 其中 :m 和 n 是权衡系数, R<sub>s</sub> 表示短期兴趣与网页情境的相关性, R<sub>l</sub> 表示长期兴趣与网页情境的相关性;

$$[0091] R_s = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Sset|} p(Cset_i | Sset_j)}{|Cset| \cdot |Sset|} R_L = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Lset|} p(Cset_i | Lset_j)}{|Cset| \cdot |Lset|}$$

[0092] 其中 :Sset 表示短期兴趣概念集合, Lset 表示长期兴趣概念集合, Cset 表示网页情境中的语境词集;

[0093] 2) 情境检索:

[0094] 此过程发生在用户提交了此次查询的查询词之后。

[0095] (1) 经权利要求 1 中步骤 4) 情境扩展模块处理获得语境词、地理位置、时间信息，利用这些信息，在权利要求 1 的步骤 3) 中构建的情境索引数据库中进行检索；

[0096] (2) 同时访问权利要求 1 中步骤 1) 建立的网页数据库，得到相匹配的网页，形式化表示为：

[0097]  $\text{Result} = \text{Search}(\text{Index}, Q, C, \text{Syn}, \text{UG}, \text{UT})$

[0098] 用 C 和 Q 或 C 和 Syn 在索引 Index 中的第一层语境索引 CI 中进行检索，用用户所处地点 UG 在第二层空间索引 GI 中进行检索，用用户最近检索时间 UT 在第三层索引时间 TI 中进行检索；在每一层检索的具体过程描述如下：

[0099] a. 在语境层检索：

[0100] 构建查询词的语境检索项，语境检索既利用消解歧义后的语境词实现精准检索，又通过查询词的同义词集实现语义相关检索，同义词与查询词具有共同的语境词，这样每个查询词  $Q_i$  产生的语境检索项表示为： $\{\langle Q_i, C_i \rangle, \langle \text{Syn}_{i1}, C_i \rangle, \dots, \langle \text{Syn}_{ik}, C_i \rangle\}$ 。

[0101] 其次，在语境索引中，通过查询词的语境检索项检索相应的语境索引项，进而在空间索引中继续检索；

[0102] b. 在空间层检索：

[0103] 借助地理学本体和地理信息系统将查询中涉及空间信息的查询词项映射为区域或坐标，并采用自然语言处理技术识别出与空间信息相关的语义信息，从而实现查询词空间敏感信息的检测；

[0104] 根据检测结果进一步判断，若查询词不具有空间敏感性，检索的内容包括语境索引项下的所有空间索引项；否则，利用用户所在位置的空间信息 UG 和查询词项涉及的空间敏感信息，确定适合空间检索的窗口大小，从而在空间索引 GI 中检索符合用户空间信息需求的相关索引项，进而在时间索引中继续检索；

[0105] c. 在时间层检索：

[0106] 对于时间不敏感查询，其检索范围为整个时间轴；对于时间敏感查询，检索范围为 [st, et] 表达的时间区间，st 为起始时间，et 为终止时间，是经过情境扩展模块处理用户提交查询词等信息分析后得到的，例如，最近 1 小时或 1 天或 1 周或 1 月或 1 年可以统一表示为过去起始时间到当前时间的时间区间。用 UT 在 B+ 树 TI 中进行检索，将落在 UT 范围内的网页作为返回结果；

[0107] 3) 情境排序：

[0108] 对于在情境检索返回的结果，依据情境相关性及网页权威性进行排序，排序之后的结果就可以提供到用户查询界面中了，情境排序的形式化表示为：

[0109]  $\text{Result}' = \text{Rank}(\text{Result}, A, R)$

[0110] 其中：A 为网页权威性，采用经典的基于网页链接的 PageRank 静态排序算法，R 为情境相关性，是用户情境表示  $UC = \langle L, S, I, Q, \text{UG}, \text{UT} \rangle$  与网页情境表示  $PC = \langle \{W, C\}, PG, PT \rangle$  匹配的结果，即： $R = \text{Match}(UC, PC)$ 。该公式基于 Q 与 W 的匹配、I 与 C 的匹配、UG 与 PG 的匹配、UT 与 PT 匹配，返回相应网页的情境相关性值；

[0111] 本发明通过实验确定参数 A 和 R 在最终排序算法中的权重，并依据该权重 对经典排序结果进行调整；

[0112] (1) 采用经典的基于网页链接的 PageRank 静态排序算法, 确定网页权威性 A 的值;

[0113] (2) 使用 BM25 概率检索模型, 通过计算网页情境中的语境词与查询意图间的相关性概率, 获得网页文档与查询词之间的概念相关性得分  $R_c$ 。即在查询词扩展所得的各上位概念, 与网页情境解析所得的各上位概念之间, 计算相关性概率。文档的分值可以匹配到各概念项概率求和;

[0114] (3) 网页情境解析的某上位概念  $C_x$  的概率, 可以由相关集合中所有扩展出  $C_x$  的词的频率相加求和得到;

[0115] (4) 通过比对 UG 与 PG 中各地理位置信息, 获得 UG 和 PG 的匹配度得分  $R_g$ , 计算与查询词匹配之后所得向量的分量进行加权求和得出, 即  $\sum g_i \times w_i$ ,  $g_i$  表示地理位置信息的向量各分量,  $w_i$  为对应的权重, 取值区间为 [0, 1] ;

[0116] (5) 索引项的时间信息得分  $R_t$ , 由网页情境时间信息是否匹配查询意图要求的信息, 如果匹配取值为 1, 否则取值为 0, 乘以权值常数 0.85 以保证其影响力, 再加上由查询时间和网页发布时间差值函数  $f(t)$  确定的网页新鲜程度得分, 其中  $f(t) = \text{DecayRate}^{(St-Pt)}$ ,  $St$  为查询时间,  $Pt$  为网页发布时间,  $\text{DecayRate}$  取值为 0.5 ;

[0117] (6) 最终由  $R_c+R_g+R_t$  得出 R 的值;

[0118] (7) 根据计算结果进行排序, 排序之后的结果就可以提供到用户查询界面中。

[0119] 与现有技术相比本发明的有益效果是:

[0120] 1. 本发明所述的基于本体的情境搜索方法给出了清晰、明确、可计算的网页情境和用户情境信息表示。根据词在网页中的上下文, 基于通用本体, 利用词义消歧技术, 确定词的情境, 对文档在语境层面上进行解析, 基于本体表示用户情境, 给出网页情境和用户情境的形式化表示。情境信息对于提高搜索性能至关重要, 情境信息必须以恰当的形式表示出来, 才能方便检索使用。而传统搜索引擎将网页内容表示为若干词汇的文本向量, 这种表示方法缺少词的语境和时间、空间信息等网页情境信息。另外, 传统搜索引擎在搜索过程中没有考虑用户兴趣和查询意图等用户情境信息。而本发明所提出的情境搜索就是要将这些情境信息融入搜索, 在情境级别上进行匹配, 提高用户对搜索服务的满意度。

[0121] 2. 本发明所述的基于本体的情境搜索方法构建的情境索引从根本上解决目前索引结构不支持情境搜索的问题, 摒弃先检索再过滤的传统做法, 实现所搜即所需, 提高效率和搜索精度。传统搜索引擎的索引结构包括网页到词的正向索引和词到网页的倒排索引, 缺少网页情境信息, 不利于用户情境的应用, 并且如果想在传统索引基础上进行个性化情境搜索, 则必须对返回结果进一步做过滤处理, 导致效率低、效果差。面对传统索引的不足和情境搜索带来的挑战, 需要将网页情境信息融入索引结构。该结构可以在不过多提升空间代价的情况下, 保证索引更新快捷方便, 满足实时响应要求, 提高检索系统的服务效果, 彻底摒弃“先搜索、再过滤”的传统检索方式。

[0122] 3. 本发明所述的基于本体的情境搜索方法提出了一个新概念, 情境扩展, 是指利用用户情境, 通过一定方法和策略, 在情境级别上对查询词进行扩展, 限定查询词所属的语境(词释义的直接上位概念), 形成查询情境  $\langle\{Q, \text{Syn}, C\}, \text{UG}, \text{UT}\rangle$ , 其中, Q 为查询词, Syn 为 Q 的同义词集, C 为 Q 的语境词, UG 为查询地点, UT 为查询时间。从而弥补用户查询信息不足的缺陷。例如, 对“苹果电脑”感兴趣的用户提交查询词“apple”, 情境扩展会限定

语境词为“电子产品”，这样返回结果中就不再出现诸如“水果”、“电影”类网页。根据用户的查询意图和查询词，利用本体和情境索引中的语境词，以及用户地理位置和时间检索区间，实现情境扩展，情境扩展与查询扩展有着本质的不同，区别在于：查询扩展中的扩展词作为查询词序列的一部分，必然出现在查询结果网页中；而情境扩展中限定的语境词一般并不出现在结果网页中，它只给出查询所属的概念范畴。通过情境扩展，可以更加有效地引导用户快速定位并得到所期望的搜索结果。

[0123] 4. 情境检索

[0124] 传统检索模型利用基于词的倒排索引处理查询，在检索过程中未涉及查询情境，不能直接提供个性化检索服务。通常用户提交的搜索词是多义的，例如：对于搜索词apple，可能有水果、电子产品、电影等多种释义，在该情况下，传统检索模型会返回很多冗余记录。与之不同的是，情境检索在情境索引的基础上，利用查询情境检索网页，直接获得满足用户需求的查询结果，无需过滤。这也是本发明突出的特点和贡献之一。

[0125] 5. 情境排序

[0126] 能否将与用户需求最相关的高质量网页排在返回结果的最前面，是衡量搜索引擎性能的重要指标。目前搜索引擎使用的排序算法主要包括词频位置加权排序算法和超链接分析排序算法。这些方法在排序过程中未涉及用户情境，忽略了搜索个体之间搜索情境的差异，导致当不同用户提交相同查询时，返回相同的搜索结果。另外，由于用户的思维状态是动态变化的，查询意图的预测不能保证完全正确，因此，需要依据用户后续的浏览行为修整返回结果，而目前的搜索引擎所使用的排序算法不能依据用户的后续行为，重新调整搜索结果。简言之，传统结果排序方法无法根据用户情境实现情境排序。本发明的情境排序既考虑了网页的权威性，又考虑了网页情境与用户情境，增加了排序的相关性。通过对用户交互行为进行分析，推测用户在当前查询会话中对多个交叠任务的偏重程度，进一步明确用户的信息需求，实现动态的搜索结果修整。

[0127] 6. 情境推荐

[0128] 与当前搜索引擎提供的导航不同，情境推荐并非是单纯结合用户历史点击、浏览行为数据向用户推送网址，而是以用户当前情境（包括用户长期兴趣、短期兴趣、当前时间、所处地点等信息）为推测依据，结合网页情境，向用户自动推荐一定时段内最新抓取的、具有一定权威性的、与用户情境相匹配的网页，实现“不搜变搜”、“不搜亦得”的个性化、智能化搜索目标。

[0129] 综上所述，本发明是针对情境的多维性、多样性、动态性和延展性等特征，在数据结构级别上重新设计搜索引擎，而不是在现有搜索引擎数据结构基础上进行增补和改进，从基础上解决情境搜索问题，从而实现情境化智能的搜索引擎。

## 附图说明

[0130] 下面结合附图对本发明作进一步的说明：

[0131] 图1是实施本发明所述的基于本体的情境搜索方法的计算机程序的各组成模块的功能、连接关系示意框图；

[0132] 图2是本发明所述的基于本体的情境搜索方法的流程框图；

[0133] 图3是本发明所述的基于本体的情境搜索方法中由网页情境解析与表示模块对

采集到的网页进行解析并表示网页情境信息的流程框图；

[0134] 图 4 是本发明所述的基于本体的情境搜索方法中由用户情景挖掘与表示模块实现用户情境挖掘与表示的流程框图；

[0135] 图 5 是本发明所述的基于本体的情境搜索方法中由情境索引处理模块设计并实现融入网页情境信息的索引结构的示意框图；

[0136] 图 6 是本发明所述的基于本体的情境搜索方法中由情境查询扩展模块通过情境扩展更加有效地引导用户快速定位并得到所期望的搜索结果的流程框图；

[0137] 图 7 是本发明所述的基于本体的情境搜索方法中由情境查询处理模块实现情境推荐、情境检索，并使用情境排序子模块对情境检索结果加工处理，最后将所得结果返回到用户的查询界面的流程框图；

[0138] 图 8 是基于本体的情境搜索方法中经过用户兴趣挖掘而得到的某用户兴趣图；

图 9 是本发明所述的基于本体的情境搜索方法中实现网页情境解析与表示以及构建情境索引过程中处理的一个中文网页实例，该网页的网址为 <http://tech.sina.com.cn/mobile/n/apple/2012-12-11/08377875411.shtml>。

### 具体实施方式

[0139] 下面结合附图对本发明作详细的描述：

[0140] 本发明所述的基于本体的情境搜索方法所要解决的技术问题是克服了现有技术的不足，基于本体提出网页情境解析与表示、用户情境挖掘与表示、情境索引处理、情境扩展、情境查询处理等情境搜索中的关键问题，提出并实现一系列面向情境搜索的新技术与新方法，可有效地解决面向情境的搜索问题，提高搜索的精准性，实现个性化搜索，达到所得即所需的服务目标，为新一代搜索引擎提供支持。其中本体是关于领域知识的概念化、形式化的明确规范，定义了组成主题领域的词汇表的基本术语及其关系，以及结合这些术语和关系来定义词汇表外延的规则。经过仔细分析，本方法采用的本体技术作为一种通用工具，特别适合情境搜索问题。本发明所使用的本体是在现有本体的基础上进一步扩充和完善后所生成的通用本体。

[0141] 参阅图 1，为实现基于本体的情境搜索方法这一目的自行编制了计算机程序，其包括有五个功能模块，即包括网页情境解析与表示模块、用户情境挖掘与表示模块、情境索引处理模块、情境查询扩展模块和情境查询处理模块，各个模块功能：

[0142] 1. 网页情境解析与表示模块

[0143] 所述的网页情境解析与表示模块分为网页情境解析和网页情境表示两部分，网页情境包括词、语境词、网页地点、更新时间等。首先进行网页采集，对采集到的网页建立 DOM 树，实现网页去噪、提取正文等网页解析操作，然后通过自然语言处理技术和通用本体进行词义消歧，实现网页的情境解析，利用命名实体识别技术获取文档内的时间和地点信息，最终将获取到的信息组织输出为网页情境表示形式。

[0144] (1) 网页情境解析：

[0145] 为了支持基于情境的检索，必须确定文档中多义实词的唯一释义，网页情境解析根据文档上下文信息确定多义实词的唯一释义，本发明采用基于本体的词义消歧技术来实现。由于词的一个释义可以存在多层上位概念，本发明中采用词的直接上位概念作为其语境。

[0146] (2) 网页情境表示：

[0147] 网页情境表示是构建情境索引的基础，也是实现情境检索的保障。传统搜索引擎将网页内容表示为文本向量，不含有词的语境和时间、空间信息。本发明提出的网页情境包括：(1) 词和语境词；(2) 网页内的地点信息；(3) 网页时间信息；包括网页创建时间、最近更新时间、网页内时间。

[0148] 本发明中利用通用本体表示网页情境，并结合函数描述，给出了其形式化表示。

[0149] 2. 用户情境挖掘与表示模块

[0150] 所述的用户情境挖掘与表示模块分为用户情境挖掘和用户情境表示两部分，负责使用本体技术给出清晰、明确、可计算的用户情境说明和描述。用户情境包括用户兴趣、查询意图以及搜索时间和地点等。通过挖掘用户兴趣、从用户连续的搜索行为中预测查询意图，同时考虑用户搜索时所处的时间和空间因素，形成用户情境表示，为情境扩展和排序提供依据。情境搜索强调以“人”为本，深层挖掘用户情境，准确定位用户需求。具体内容包括：

[0151] (1) 用户情境挖掘：

[0152] 用户情境包括用户兴趣、查询意图以及搜索时间和地点，其中，提交查询的时间和地点易于获得，而通过用户搜索日志挖掘用户兴趣，并在此基础上推断查询意图，则是用户情境发现的难点。用户兴趣包括长期兴趣和短期兴趣。长期兴趣描述了用户搜索需求的长期特征，通常与爱好、专业或职业等有关，具有相对稳定的特性；短期兴趣刻画了用户短时间内关注的内容，是对短期检索需求的反映，具有随时间漂移的特性。本发明以用户兴趣为基础，从用户连续的搜索行为中预测查询意图，实现零负荷、高精度的用户情境获取，这对于情境搜索来说至关重要。由于用户行为变量和意图标签数量庞大，在用户搜索过程中，直接模型化表示用户行为动态或用户意图动态是非常困难的。鉴于贝叶斯网络具有以下特点：

[0153] ①可以处理不完整和带有噪声的数据集；

[0154] ②用图形的方法描述数据间的相互关系，语义清晰，可理解性强；

[0155] ③易将先验知识和概率相结合，具有因果和概率性语义。

[0156] 本发明基于贝叶斯网络推断用户搜索过程中的查询意图。

[0157] (2) 用户情境表示：

[0158] 系统充分理解用户情境是实现情境搜索服务目标的重要前提，因此给出清晰、明确、可计算的用户情境说明和描述是十分必要的。目前信息表示的主要方法有词袋、向量和图，但均不足以应对用户情境表示面临的各种挑战，本发明采用本体技术组织和描述用户情境。

[0159] 3. 情境索引处理模块

[0160] 所述的情境索引处理模块是在网页情境解析与表示的基础上进行的，负责调用网页情境解析模块，并根据分析结果，设计融入了语境词和时间、空间信息的情境正向索引结构，并将语境索引与时空索引有机地结合在一起，建立真正直接支持情境检索的索引结构。传统搜索引擎的索引结构包括网页到词的正向索引和词到网页的倒排索引，如果在该索引结构的基础上实现个性化情境搜索，则必须对返回结果进一步过滤处理。面对传统索引的不足和情境搜索带来的挑战，需要将网页情境信息融入索引结构。该结构可以在不过多提

升空间代价的情况下,设计情境正向索引和情境倒排索引。

[0161] (1) 情境正向索引 :

[0162] 传统搜索引擎的正向索引记录了网页到词的对应关系。这种正向索引结构简单,在其基础上可以方便快捷地构建词到网页的倒排索引,但缺少语境信息。本发明利用网页情境表示,在传统正向索引的基础上,设计融入了语境词和时间、空间信息的情境正向索引结构。

[0163] (2) 情境倒排索引 :

[0164] 传统搜索引擎的倒排索引记录了词到网页的对应关系。这种倒排索引的并不直接支持情境检索。本发明利用情境正向索引,设计了〈词,语境词〉到网页列表的语境索引结构,并借助时空数据库技术,建立基于 R\* 树的空间索引和基于 B+ 树的时间索引,最后将语境索引与时空索引有机地结合在一起,得到真正直接支持情境检索的索引结构。

[0165] 4. 情境查询扩展模块

[0166] 情境扩展是本发明中提出的一个新概念,该模块根据用户提供的查询词和用户情境挖掘与表示模块中表示出来的用户情境信息相结合,在通用本体中选取恰当的语境词作为情境扩展提供给用户,完善用户查询,准确概括和描述用户信息和需求,从而弥补用户查询信息不足的缺陷。对于多个语境词要按匹配情况进行排序,用户选定扩展的语境词后,本发明可以根据需要更新该用户所对应的兴趣和查询意图。值得注意的是“情境扩展”与当前搜索引擎,如百度、google 等,提供的“查询扩展”存在本质不同,本发明中提出的“情境扩展”所限定的语境词一般并不需要出现在返回的结果网页中,它只给出查询所属的概念范畴,而“查询扩展”所扩展的内容作为查询词序列的一部分,必须出现在返回的结果网页中。

[0167] 5. 情境查询处理模块

[0168] 所述的情境查询处理模块包括情境推荐、情境检索和情境排序三个子模块:

[0169] 1) 情境推荐子模块负责将网页数据库中存储的最新的网页,根据网页情境与用户情境的匹配程度,按照匹配分值高的网页主动推荐(推送)给用户,达到“不搜亦得”的智能化检索目标,这与当前搜索引擎提供的导航不同,情境推荐并非单纯结合用户点击浏览行为数据向用户推送网址,而是以用户当前情境(包括用户的长期兴趣、短期兴趣、当前时间、所处地点等信息)为推测依据,结合网页情境,向用户自动推荐一定时段内最新抓取的、具有一定权威性的、与用户情境相匹配的网页;

[0170] 2) 情境检索子模块发生在用户提交查询词之后,负责在情境索引的基础上,利用经情境扩展获得的查询情境信息检索网页,直接获得满足用户需求的查询结果,无需过滤。在实际检索中,检索提问涉及的查询词往往不只一个,而同一个查询词通常又涉及多个同义词或相关词,查询所处的用户情境更是千差万别,如何有效查找到符合用户情境的网页,是情境检索子模块的主要任务;

[0171] 3) 情境排序子模块负责对情境检索返回的结果,结合用户情境和网页情境增加排序的相关性,并且依据情境相关性(语境相关性、网页权威性、地理位置、时间等)进行排序,目前搜索引擎使用的排序算法主要包括词频位置加权 排序算法和链接分析排序算法,排序过程未涉及用户情境。

[0172] 参阅图 2,本发明所述的基于本体的情境搜索方法的步骤如下:

[0173] 1. 参阅图 3,由网页情境解析与表示模块对采集到的网页进行解析并表示网页情境信息,步骤如下:

[0174] 1) 使用网络爬虫 (Crawler) 采集网页并存入网页数据库;

[0175] 2) 建立 DOM 树进行网页解析,并使用实时搜索爬虫提取网页最近更新时间;

[0176] 3) 实现网页情境解析:

[0177] (1) 如果网页是中文网页,则对网页进行分词;

[0178] (2) 利用通用本体和自然语言处理技术构建语义关联图,从而对词或术语进行词义标注。

[0179] 构建语义关联图的过程为:

[0180] 歧义词 W 的往往有多个意义,每个意义就是一个义项,利用通用本体,以歧义词 W 的各义项为中心,构建半径为 N(本发明中取值为 3) 的包含如下关系的语义关联图:

[0181] a. 义项描述关系 (is-gloss);

[0182] b. 实例关系 (is-sample);

[0183] c. 结构化语义关系:上位、下位、整体、部分、同义、反义、蕴含、因果、近似、附属等;

[0184] d. 领域属性关系 (is-domain);

[0185] 语义关联图以边为单位用三元组方式存储,即每一条语义关系作为一个记录存储,语义关系是有向关系,表示方法为:{left, right, Relation Name},例如:{bus#1, public transport#1, kind-of}。

[0186] (3) 确定语义关联图中语义关系权重;语义关系权重的确定采用如下公式:

$$SR(Sen_1, Sen_2) = \max_r - \frac{(\max_r - \min_r)}{n_r(Sen_1)}$$

[0188] 其中: $\max_r$ 、 $\min_r$  是赋予关系 r 的最大权重与最小权重,  $n_r(Sen_1)$  是在语义关联图中从词义  $Sen_1$  出发的关系 r 的边的个数。

[0189] (4) 计算歧义词 W 的每个义项和上下文词集的关联度,公式为:

$$\sum_{j=1}^{|CS|} WCR(WS_i, CS_j)$$

[0191] 其中: $WS_i$  为歧义词 W 的第 i 个义项, CS 为满足以下条件的义项集合:①出现在语义关联图中,②为上下文词集中词的一个义项, $WCR(WS_i, CS_j)$  为基于语义关联图和语义关系权重计算义项  $WS_i$  和  $CS_j$  的关联度函数;

[0192] (5) 选取关联度最大的义项作为词 W 在上下文中的释义,并提取释义的直接上位概念作为语境词,获取 W 的语境,从而实现文档的情境解析。

[0193] 4) 利用命名实体识别技术获取文档内的时间与地点信息,并将文档表示为网页情境形式:

[0194]  $PC = \langle \{W, C\}, PG, PT \rangle$

[0195] 其中:W 为词或术语,C 为语境词,PG 为网页内的地点信息,PT 为网页时间信息(包含创建时间、更新时间、网页内时间)。

[0196] 语义标注是网页情境解析的关键问题,本发明采用 Senseval 会议提供的测试语料集进行词义标注实验,该会议为参加者提供了统一的评测语料、词义目录和评测标准,已

经成为了词义消歧领域事实上的测试标准。

[0197] 2. 参阅图4,由用户情境挖掘与表示模块实现用户情境挖掘与表示并和第1步骤同时进行,步骤如下:

[0198] 1) 用户兴趣挖掘:

[0199] 借助通用本体,通过分析历史数据获得用户长期和短期兴趣,建立用户兴趣模型;

[0200] (1) 对用户搜索日志进行挖掘,获取用户点击的网页类别;

[0201] (2) 量化用户点击各类网页的会话(session)数和频度;

[0202] (3) 设定阈值,依据频度,提取用户长期兴趣概念集  $C_L$  和短期兴趣概念集  $C_S$ ;

[0203] 用户长期兴趣与短期兴趣分别用通用本体的局部来表示,并可动态漂移。针对用户兴趣的漂移性,用户的长期兴趣和短期兴趣漂移表示为:

[0204]  $L' = \text{long\_term\_interest\_drift}(L, S, H)$

[0205]  $S' = \text{short\_term\_interest\_drift}(S, L, H)$

[0206] 其中:L为用户长期兴趣,S为用户短期兴趣,H为查询历史数据,L'和S'分别为用户兴趣增量学习过程获得的长期兴趣和短期兴趣。

[0207] (4) 以用户兴趣为基础,基于通用本体,提取  $C_L$  和  $C_S$  中的上下位词汇链,将各词汇链衔接在一起,即得用户兴趣模型。

[0208] 词汇链是由一组概念通过语义关系连接形成的概念语义链,它表示了概念在语义上的内聚关系。以不同语义关系连接概念,可形成不同的词汇链。在此,本发明以上下位关系连接  $C_L$  和  $C_S$  中的概念,构建上下位词汇链。依次合并  $C_L$  中蕴含的上下位词汇链,构造长期兴趣模型。同上,依次合并  $C_S$  中蕴含的上下位词汇链,构造短期兴趣模型。

[0209] 2) 查询意图预测:

[0210] 用户当前的查询意图,可能和上次查询意图相同(如:当前查询是重复查询),也可能属于用户长期关注的领域,亦有可能属于用户最近关注的领域。

[0211] (1) 根据用户本次的查询词,从用户连续的搜索行为中预测查询意图,并 实现用户情境获取确定用户的查询意图。采用动态贝叶斯网络推断当前用户查询的搜索意图,形式化表示为:

[0212]  $I = \text{Intent\_infer}(L, S, P, Q) = \langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$

[0213] 其中,  $p_1, p_2, \dots, p_n$  分别为查询子意图  $I_1, I_2, \dots, I_n$  的概率(可信度),I为用户的查询意图,L为用户的长期兴趣,S为用户的短期兴趣,P为点击网页,Q为查询,Intent\_infer 为基于动态贝叶斯网的查询意图推测函数。

[0214] (2) 通过动态贝叶斯网络,利用搜索系统的先验知识,计算用户当前查询意图的后验概率分布。本发明根据已有的经验知识,定义了①用户的查询意图I、②用户的长期兴趣L、③用户的短期兴趣S、④点击网页P与⑤查询Q五个数据变量,并找出了数据变量之间的因果关系,获得动态贝叶斯网络。

[0215] 查询意图预测涉及的主要概率计算公式如下:

[0216]

$$p(I_t | I_{t-1}, L_t, S_t, P_{t-1}, Q_t) = \frac{p(I_t, I_{t-1}, L_t, S_t, P_{t-1}, Q_t)}{p(I_{t-1}, L_t, S_t, P_{t-1}, Q_t)} = \frac{p(I_t)p(I_{t-1}|I_t)p(L_t|I_t)p(S_t|I_t)p(P_{t-1}|I_t)p(Q_t|I_t)}{p(I_{t-1})p(L_t)p(S_t)p(P_{t-1})p(Q_t)}$$

[0217] 其中 : $I_t$  表示第 t 次查询的意图,  $I_{t-1}$  表示第 t-1 次查询的意图,  $L_t$  表示第 t 次查询时的用户长期兴趣,  $S_t$  表示第 t 次查询时的用户短期兴趣,  $P_{t-1}$  表示第 t-1 次查询时的点击的网页,  $Q_t$  表示第 t 次查询。

[0218] 3) 基于步骤 1) 和步骤 2) 的计算, 使用本体技术给出清晰、明确、可计算的用户情境说明和描述。用户情境包括用户兴趣、查询意图以及搜索时间和地点。可以将用户情境表示为 :

[0219]  $UC = \langle L, S, I, Q, UG, UT \rangle$

[0220] 其中 : $L$  为用户长期兴趣,  $S$  为用户短期兴趣,  $I$  为当前查询意图,  $Q$  为查询词,  $UG$  为用户地理位置,  $UT$  为用户查询时间。

[0221] 3. 由情境索引处理模块设计融入网页情境信息的索引结构, 步骤如下 :

[0222] 参阅图 5, 在实现了网页情境解析和表示之后, 根据步骤 1 中获得的网页情境信息进行情境正向索引和情境倒排索引的创建, 由所述的情境索引处理模块设计并实现融入网页情境信息的索引结构, 该结构可以在不过多提升空间代价的情况下, 保证索引更新快捷方便, 满足实时响应要求, 提高检索系统的服务效果, 可以摒弃“先搜索、再过滤”的传统检索方式。

[0223] 1) 构建情境正向索引 :

[0224] 根据网页情境解析的结果, 构造文档到〈词, 语境词〉以及时间与空间信息之间的索引, 同时附着索引项的相应特征信息, 包括词在网页中位置、词频、权重等。本发明将情境正向索引表示为 :

[0225]  $docid \rightarrow \langle \{ \langle wordid, context \rangle, \{pos\}, freq, weight \}, \{Time\}, \{Geo\} \rangle$

[0226] 其中 docid 表示文档号, wordid 表示词号, context 表示语境词, pos 表示词在网页中的位置, freq 表示词在网页中的词频, weight 表示词在网页中的权重, Time 表示时间信息, Geo 表示空间信息。

[0227] 情境正向索引主要分为三个部分 : 语境信息、时间信息和空间信息, 来自于网页情境解析与表示模块处理后得到的结果信息, 每一部分信息描述如下 :

[0228] a. 语境信息 :

[0229] 表达式中的  $\{ \langle wordid, context \rangle, \{pos\}, freq, weight \}$  部分合称为语境信息, 是在网页情境解析后, 将词和语境词绑定供检索匹配时使用, 一个词的语境词用该词在本体中的直接上位概念表示, 由于语境词本身可能也是检索词, 因而在本体和倒排索引中也有其上位概念, 如此可以找到一串上位词 (语境词), 称作“词汇链”, 供检索匹配时使用。本发明中将 wordid 与 context 绑定, 以方便构造倒排索引 ;

[0230] b. 时间信息 : 表达式中的 {Time} 部分, 由网页创建时间、更新时间和网页内容时间等构成, 并以标准日期、时间格式表示 ;

[0231] c. 空间信息 : 表达式中的 {Geo} 部分, 表示网页内容中所涉及的地点信息, 并以经度和纬度形式表示。

[0232] 这样, 情境正向索引实现了将网页解析后各种情境信息表示成索引项的形式, 以便在下面的步骤 2) 中构建情境倒排索引。

[0233] 2) 构建情境倒排索引 :

[0234] 情境倒排索引是在传统搜索引擎倒排索引技术基础上, 融入了语境信息、时间信

息和空间信息,得到真正直接支持情境检索的索引结构。

[0235] (1) 首先利用情境正向索引,构造〈词,语境词〉到文档集合的索引,即:

[0236]  $\langle \text{wordid}, \text{context} \rangle \rightarrow \{\langle \text{docid} \rangle\}$

[0237] 本发明中将 wordid 与 context 绑定,以方便情境检索与情境扩展。注意在检索时,用检索词对倒排索引中的 wordid 进行匹配,将用户情境与 context 进行匹配。

[0238] 与通常搜索引擎索引结构相比,本发明设计并实现的情境索引增加了语境词 (context)。由于一个词的释义一般不超过 5 个,多在 2-3 个之间,只考虑语境词的索引存储规模的增加不会超过 5 倍,因此,本发明设计的情境索引是非常简洁而有效的。

[0239] (2) 为支持对时间和空间敏感的查询,需要借鉴时空数据库的索引结构对上述提出的索引进行进一步的改进,空间索引 GI 用 R\* 树构建,时间索引 TI 用 B+ 树构建,语境索引 CI 用倒排形式构建,对 CI 中的 docid 列表先按地理位置,再按时间进行索引。三层索引表示为 Index = CI → GI → TI,注意 CI 中的 doclist(文档序列)已被 GI 和 TI 进一步索引。

[0240] 4. 由情境扩展模块实现情境扩展并形成查询情境,步骤如下:

[0241] 参阅图 6,根据步骤 2 即用户情境挖掘与表示模块中得到的用户情境信息,利用通用本体实现情境扩展,本发明提出的情境扩展是一个新概念,是指利用用户情境,通过一定方法和策略,在情境级别上对查询词进行扩展,限定查询词所属的语境,形成查询情境,从而弥补用户查询信息不足的缺陷。例如,当对“苹果电脑”感兴趣的用户提交查询词“apple”时,情境扩展会限定语境词为“电子产品”,这样返回结果中就不再出现诸如“水果”、“电影”类网页。情境扩展与查询扩展有着本质的不同,区别在于:查询扩展中的扩展词作为查询词序列的一部分,必然出现在查询结果网页中;而情境扩展中限定的语境词一般并不需要出现在结果网页中,它只给出查询所属的概念范畴。通过情境扩展,可以更加有效地引导用户快速定位并得到所期望的搜索结果。

[0242] 本发明利用通用本体选取恰当的语境词作为情境扩展提供给用户,由于一个术语在通用本体中可能有多个语境词与之对应,因此需要依据用户情境信息选取最佳匹配的语境词。本发明中,利用通用本体结合查询词和步骤 2 中得到的查询意图获得相应语境词,实现情境扩展,形式化表示为:

[0243]  $\text{SearchContext} = \text{ContextExpansion}(Q, O, I, UG, UT)$

[0244] 其中:Q 为查询词, O 为通用本体, I 为查询意图, UG 为查询地点, UT 为查询时间。在绝大多数情况下,本发明推荐的语境词是恰当的,但当用户意图发生突变时,最佳匹配的语境词不一定是本次搜索所期望的,此时本发明将提供次佳匹配的语境词。因此,需要将多个语境词按匹配情况进行排序。

[0245] 1) 根据用户提供的查询词和在步骤 2 中用户情境挖掘与表示模块所获得的用户情境信息,使用通用本体提取查询 Q 的若干释义  $\langle QS_1, QS_2, \dots, QS_m \rangle$  及相应的若干语境  $\langle C_1, C_2, \dots, C_m \rangle$ 。

[0246] 2) 在步骤 2 中用户情境挖掘与表示模块所获得的查询意图  $\langle I_1(p_1), I_2(p_2), \dots, I_n(p_n) \rangle$  的基础上,利用朴素贝叶斯模型计算语境词  $C_i$  的后验概率  $p(C_i | I_1, I_2, \dots, I_n)$ ,将多个语境词按匹配情况进行降序排序,并按照该值的降序向用户推荐相应语境的情境扩展,其中最大值所对应的语境词为最佳语境词,其次为次佳语境词。

[0247] 3) 若用户选择某情境扩展，则取相应语境为最终确认结果，否则自动选取  $p(C_i | I_1, I_2, \dots, I_n)$  值最大的语境为最终确认结果，且以次佳匹配的语境词为备选。用户选定扩展的语境词后，如果用户选定的语境词与本发明提供的最佳语境词不一致，需要更新该用户的兴趣和查询意图。

[0248] 4) 最后获取查询词 Q 的同义词集 Syn，查询情境形式化表示为： $\langle\langle Q, Syn, C \rangle\rangle, UG, UT$ ，其中 :Q 为查询词，Syn 为 Q 的同义词集，C 为 Q 的语境词，UG 为查询地点，UT 为查询时间。

[0249] 5. 由情境查询处理模块实现情境查询处理，步骤如下：

[0250] 使用步骤 1 中网页情境挖掘和表示模块所得到的网页数据库，以及步骤 3 中情境索引处理模块所创建的索引数据库，结合步骤 3 中获得的用户情境表示，并调用步骤 4 情境扩展模块得到的查询扩展结果，实现情境推荐、情境检索，并由情境排序子模块对情境检索结果加工处理，最后将所得结果返回到用户的查询界面。

[0251] 1) 情境推荐：

[0252] 此步骤发生在用户打开查询界面但是没有进行查询操作时，对于经步骤 1 网页采集存储在网页数据库中网页，根据网页情境与用户情境的匹配程度及网页权威性，以合理的顺序向用户主动推荐网页，达到“不搜亦得”的智能化检索目标。情境推荐的形式化表示为：

[0253]  $Result = Recommend(Pages, UT, UG, A, R_{SL})$

[0254] 其中 :Pages 表示新抓取的网页集合，UT 表示用户最近查询时间，UG 表示用户所处地点，A 表示网页权威性， $R_{SL}$  表示用户兴趣与网页情境的相关性。

[0255] 情境推荐形式化表示中的各项的计算方法如下：

[0256] a. 网页权威性 A 的计算采用经典的基于网页链接的 PageRank 静态排序算法。

[0257] b. 用户兴趣与网页情境的相关性  $R_{SL} = m \times R_s + n \times R_l$ 。其中 :m 和 n 是权衡系数， $R_s$  表示短期兴趣与网页情境的相关性， $R_l$  表示长期兴趣与网页情境的相关性。

$$[0258] R_s = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Sset|} p(Cset_i | Sset_j)}{|Cset| \cdot |Sset|} R_l = \frac{\sum_{i=1}^{|Cset|} \sum_{j=1}^{|Lset|} p(Cset_i | Lset_j)}{|Cset| \cdot |Lset|}$$

[0259] 其中 :Sset 表示短期兴趣概念集合，Lset 表示长期兴趣概念集合，Cset 表示网页情境中的语境词集。

[0260] 2) 情境检索：

[0261] 此过程发生在用户提交了此次查询的查询词之后。

[0262] (1) 经步骤 4 情境扩展模块处理获得语境词、地理位置、时间等信息，利用这些信息，在步骤 3 中构建的情境索引数据库中进行检索。

[0263] (2) 同时访问步骤 1 建立的网页数据库，得到相匹配的网页，形式化表示为：

[0264]  $Result = Search(Index, Q, C, Syn, UG, UT)$

[0265] 用 C 和 Q 或 C 和 Syn 在索引 Index 中的第一层语境索引 CI 中进行检索，用用户所处地点 UG 在第二层空间索引 GI 中进行检索，用用户最近检索时间 UT 在第三层索引时间 TI 中进行检索。由于采用了恰当的数据结构，保证快速定位检索信息。

[0266] 在每一层检索的具体过程描述如下：

[0267] a. 在语境层检索：

[0268] 首先，构建查询词的语境检索项，语境检索既可以利用消解歧义后的语境词实现精准检索，又可以通过查询词的同义词集实现语义相关检索，同义词与查询词具有共同的语境词，这样每个查询词  $Q_i$  产生的语境检索项表示为： $\{ \langle Q_i, C_i \rangle, \langle Syn_{i1}, C_i \rangle, \dots, \langle Syn_{ik}, C_i \rangle \}$ 。

[0269] 其次，在语境索引中，通过查询词的语境检索项检索相应的语境索引项，进而在空间索引中继续检索。

[0270] b. 在空间层检索：

[0271] 首先，借助地理学本体和地理信息系统将查询中涉及空间信息的查询词项映射为区域或坐标，并采用自然语言处理技术识别出与空间信息相关的语义信息，从而实现查询词空间敏感信息的检测。

[0272] 然后，根据检测结果进一步判断，若查询词不具有空间敏感性，检索的内容包括语境索引项下的所有空间索引项；否则，利用用户所在位置的空间信息  $UG$  和查询词项涉及的空间敏感信息，确定适合空间检索的窗口大小，从而在空间索引  $GI$  中检索符合用户空间信息需求的相关索引项，进而在时间索引中继续检索。

[0273] c. 在时间层检索：

[0274] 对于时间不敏感查询，其检索范围为整个时间轴；对于时间敏感查询，检索范围为  $[st, et]$  表达的时间区间， $st$  为起始时间， $et$  为终止时间，是经过情境扩展模块处理用户提交查询词等信息分析后得到的，例如，最近 1 小时或 1 天或 1 周或 1 月或 1 年可以统一表示为过去起始时间到当前时间的时间区间。用  $UT$  在  $B^+$  树  $TI$  中进行检索，将落在  $UT$  范围内的网页作为返回结果。

[0275] 3) 情境排序：

[0276] 此步骤发生在经情境检索取得返回结果之后，对于情境检索返回的结果，依据情境相关性（语境相关性、地理位置、时间等）及网页权威性进行排序，排序之后的结果就可以提供到用户查询界面中了，情境排序的形式化表示为：

[0277]  $Result' = Rank(Result, A, R)$

[0278] 其中： $A$  为网页权威性，采用经典的基于网页链接的 PageRank 静态排序算法， $R$  为情境相关性，是用户情境表示  $UC = \langle L, S, I, Q, UG, UT \rangle$  与网页情境表示  $PC = \langle \{W, C\}, PG, PT \rangle$  匹配的结果，即： $R = Match(UC, PC)$ 。该公式基于  $Q$  与  $W$  的匹配、 $I$  与  $C$  的匹配、 $UG$  与  $PG$  的匹配、 $UT$  与  $PT$  匹配，返回相应网页的情境相关性值。

[0279] 本发明通过实验确定参数  $A$  和  $R$  在最终排序算法中的权重，并依据该权重对经典排序结果进行调整。

[0280] (1) 采用经典的基于网页链接的 PageRank 静态排序算法，确定网页权威性  $A$  的值；

[0281] (2) 使用 BM25 概率检索模型，通过计算网页情境中的语境词与查询意图间的相关性概率，获得网页文档与查询词之间的概念相关性得分  $Rc$ 。即在查询词扩展所得的各上位概念，与网页情境解析所得的各上位概念之间，计算相关性概率。文档的分值可以匹配到各概念项概率求和。

[0282] (3) 网页情境解析的某上位概念  $C_x$  的概率, 可以由相关集合中所有扩展出  $C_x$  的词的频率相加求和得到。

[0283] (4) 通过比对 UG 与 PG 中各地理位置信息, 获得 UG 和 PG 的匹配度得分  $Rg$ , 计算与查询词匹配之后所得向量的分量进行加权求和得出, 即  $\sum g_i \times w_i$ ,  $g_i$  表示地理位置信息的向量各分量,  $w_i$  为对应的权重, 取值区间为 [0, 1]。

[0284] (5) 索引项的时间信息得分  $Rt$ , 由网页情境时间信息是否匹配查询意图要求的信息, 如果匹配取值为 1, 否则取值为 0, 乘以权值常数 0.85 以保证其影响力, 再加上由查询时间和网页发布时间差值函数  $f(t)$  确定的网页新鲜程度得分, 其中  $f(t) = DecayRate^{(St-Pt)}$ ,  $St$  为查询时间,  $Pt$  为网页发布时间,  $DecayRate$  取值为 0.5。

[0285] (6) 最终由  $Rc+Rg+Rt$  得出  $R$  的值。

[0286] (7) 根据计算结果进行排序, 排序之后的结果就可以提供到用户查询界面中。

[0287] 由于用户的思维状态是动态变化的, 查询意图的预测不能保证完全正确, 因此, 需要依据用户后续的浏览行为, 调整结果顺序, 或对次佳情境扩展所对应的检索结果执行情境排序操作。依据用户后续的浏览行为调整结果顺序的形式化表示为:

[0288] Result" = Adjust(Result', Behavior)

[0289] 其中 :Behavior 为用户后续的浏览行为。

[0290] 实施例:

[0291] 参阅图 2, 本发明所述的基于本体的情境搜索方法的步骤如下:

[0292] 1. 参阅图 3, 由网页情境解析与表示模块对采集到的网页进行解析并表示网页情境信息, 步骤如下:

[0293] 1) 使用网络爬虫 (Crawler) 采集网页并存入网页数据库;

[0294] 2) 建立 DOM 树进行网页解析;

[0295] 3) 实现网页情境解析:

[0296] (1) 以网址为

[0297] <http://tech.sina.com/mobile/n/apple/2012-12-11/08377875411.shtml> 的中文网页为例, 如图 9 中网页所示。

[0298]

[0299] 需要对该网页进行分词, 结果如下:

[0300] Jefferies/x 分析师/n Peter Misek/x 昨日/t 称/v, 苹果/n 下一代/niPhone/x 明年/t 六月/t 发布/v, 拥有/v 多种/m 机身颜色/n, 提升/v 摄像头/n, 支持/v NFC/x 功能/n。

[0301] 分析师/n Peter Misek/x 认为/v iPhone/x 5S/x 拥有/v 6-8/m 种 q 不同/a 机身颜色/n, 而/cc 这/rzv 似乎/d 有/vyou 意向/n iPod/x touch/x 靠拢/v。此前/t 苹果/n i Phone/x 除了/p 黑白/n 两种/m 配色/v, 从未有过/v1 其它/rz 任何/rz 颜色/n。除此之外/c, iPhone/x 5S/x 将/d 采用/v IGZO/x 超级/b 视网膜/n 屏幕/n, 新增/v 128GB/x 机身/n 容量/n, 拥有/v 更好/a 待机/vn, 支持/v NFC/x 功能/n。Peter Mi sek/x 称/v iPhone/x 5S/x 面世/v i 时间/n 明年/t 六月/t 或/c 七月/t。

[0302] 而/cc 针对/p iPhone/x 6/vg, 分析师/n 称/v 已经/d 有/vyou 了/y 多/m 部/n 原型机/n。iPhone/x 6/vg 原型机/n 拥有/v 4.8/m 英寸/q 屏幕/n, 采用/v IGZO/x

材质 /n Retina/x+/q 超级 /b 视网膜 /n 屏幕 /n, 并且 /c 搭载 /v A7/x 四核 /n 处理器 /n, 取消 /v 实体 /n HOME/x 按键 /n。

[0303] 此前 /t 台湾 /ns DigiTimes/x 曾 /d 报道 /v 称 /v, iPhone/x 5S/x 将 /d 在 /p 本月 /t 进行 /vn 小规模 /d 试产 /v, 预计 /v 5-10/m 万 /m 台 /n, 2013 年 /t 第一季度 /n 推出 /v。并 /d 有 /v you 消息 /n 称 /v, 由于 /p iPhone 5/x 采用 /v incell/x 面板 /n 良 /d 率 /v 不佳 /a, 苹果 /n 很有可能 /n 下一代 /niPhone/x 中 /f 采用 /v OGS/x 触控 /v 面板 /n, 明年 /t 上半年 /t 正式 /ad 推出 /v。(琪 /ng 欣 /vg)

[0304] (2) 利用通用本体和语义关联图对词或术语进行消歧;

[0305] (3) 确定语义关联图中语义关系权重;语义关系权重的确定采用如下公式:

$$[0306] SR(Sem_1, Sem_2) = \max_r - \frac{(\max_r - \min_r)}{n_r(Sem_1)}$$

[0307] 其中 : $\max_r$ 、 $\min_r$  是赋予关系 r 的最大权重与最小权重, 本发明中分别取  $\max_r$ 、 $\min_r$  的值为 1 和 0 ; $n_r(Sem_1)$  是在语义关联图中从词义  $Sen_1$  出发的关系 r 的边的个数。关系 r 包括①义项描述关系 (is-gloss); ②实例关系 (is-sample); ③结构化语义关系:上位、下位、整体、部分、同义、反义、蕴含、因果、近似、附属等; ④领域属性关系 (is-domain)。

[0308] (4) 计算歧义词“苹果”的每个义项和上下文词集的关联度, 公式为:

$$[0309] \sum_{j=1}^{|CS|} WCR(WS_i, CS_j)$$

[0310] 其中,  $WS_i$  为歧义词 W 的第 i 个义项, 歧义词“苹果”的义项分别为:

[0311]  $WS_1$ : 苹果, 一种常见的水果。

[0312]  $WS_2$ : 苹果公司, 一家著名的电子产品生产厂商, 原名苹果电脑。

[0313]  $WS_3$ : 苹果日报, 一份由香港上市公司壹传媒集团旗下的正体中文报纸。

[0314]  $WS_4$ : 苹果 (电影), 一部在 2007 年上映的中国电影。

[0315]  $WS_5$ : 苹果 (产品), i phone、ipad 等。

[0316] CS 为满足以下条件的义项集合:①出现在语义关联图中②为上下文词集中词的一个义项;

[0317] (5) 经过 (4) 中的关联度计算, 可以得出关联度最大的义项为  $WS_5$ , 将其作为词“苹果”在上下文中的释义, 并提取该释义的直接上位概念“电子产品”作为语境词。

[0318] 4) 利用命名实体识别技术获取文档内的时间与地点信息, 并将文档表示为网页情境形式:

[0319] 图 9 中所示的网页可以表示为以下的情境形式, 网页 PC=<{< 苹果 , Dp03 >, < 下一代 , Bc05 >, < iPhone , Wd05 >, < 明年 , Ca02 >, < 六月 , Ca04 >, < 发布 , Hc11 > ... , < 明年 , Ca02 >, < 上半年 , Ca02 >, < 正式 , Ed21 >, < 推出 , Hd29 > } , { 台湾 } , { 明年六月 , 七月 , 上半年 } > , 其中 Dp03、Bc05 等代表词所对应的语境词在通用本体中的编号。

[0320] 2. 参阅图 4, 由用户情境挖掘与表示模块实现用户情境挖掘与表示并和第 1 步骤同时进行, 步骤如下:

[0321] 1) 用户兴趣挖掘:

[0322] (1) 对某一指定用户两个月内的搜索日志进行挖掘, 经过分析, 获取该用户点击的网页的主题向量描述, 向量分量分属于不同的网页类别, 包括: 数码产品、体育、新闻、政治

等。

[0323] 对该用户所点击的网页类别出现的频度和持续的时间进行量化, 设定阈值, 依据频度, 提取用户短期兴趣概念集  $C_s$ , 并进一步学习得出长期兴趣概念集  $C_L$ 。

[0324] 短期兴趣概念集  $C_s$  为 { 政治, 选举, 总统, …, 美国 }。

[0325] 长期兴趣概念集  $C_L$  为 { linux, 编程工具, 编程语言, 搜索引擎, 社会网, 手机, 计算机, …, 篮球, 球类比赛, 职业比赛 }。

[0326] (2) 以用户兴趣为基础, 基于通用本体, 提取  $C_L$  和  $C_s$  中的上下位词汇链, 将各词汇链衔接在一起, 即得用户兴趣模型。

[0327] 参阅图 8, 图中为经过用户兴趣挖掘而得到的某用户兴趣模型图,  $O_1$ ,  $O_2$  是用户长期兴趣,  $O_3$  是该用户的短期兴趣。由图可见, 该用户的长期兴趣主要为电脑、手机等电子产品和体育方面。通过从本体中揭示它们之间的相互关系, 可以推断用户的兴趣倾向不仅局限于电脑和手机, 对与电子设备有关的其他产品, 如数码相机、psp 等方面的网页和信息, 该用户亦有可能存在兴趣。另外, 用户短期内对奥巴马, 美国大选等新闻事件的兴趣并未持续很久, 在后续探测过程中, 其对应的网页出现次数较少, 不能满足成为长期兴趣的阈值要求, 因此只能作为用户的短期兴趣。

[0328] 2) 查询意图预测 :

[0329] 例如 : 用户本次的检索词“苹果”, 根据用户兴趣的挖掘, 得知该用户的长期兴趣如图 8 中所示的  $O_1$  和  $O_2$ , 短期兴趣如图 8 中的  $O_3$  所示, 而采用动态贝叶斯网络推断当前用户查询的搜索意图, 形式化表示为 :

[0330]  $I = \text{Intent\_infer}(\{O_1, O_2\}, O_3, P, \text{苹果}) = \langle \text{电子产品} (p_1), \text{公司} (p_2), \text{电影} (p_3), \text{水果} (p_4) \dots \rangle$ 。

[0331] 从用户连续的搜索行为和主要概率计算公式 :

[0332]

$$p(I_t | I_{t-1}, L_t, S_t, P_{t-1}, Q_t) = \frac{p(I_t, I_{t-1}, L_t, S_t, P_{t-1}, Q_t)}{p(I_{t-1}, L_t, S_t, P_{t-1}, Q_t)} = \frac{p(I_t)p(I_{t-1}|I_t)p(L_t|I_t)p(S_t|I_t)p(P_{t-1}|I_t)p(Q_t|I_t)}{p(I_{t-1})p(L_t)p(S_t)p(P_{t-1})p(Q_t)}$$

[0333] 其中 :  $I_t$  表示第  $t$  次查询的意图,  $I_{t-1}$  表示第  $t-1$  次查询的意图,  $L_t$  表示第  $t$  次查询时的用户长期兴趣,  $S_t$  表示第  $t$  次查询时的用户短期兴趣,  $P_{t-1}$  表示第  $t-1$  次查询时的点击的网页,  $Q_t$  表示第  $t$  次查询。

[0334] 计算并推断出用户查询的搜索意图 :

[0335]  $I = \langle \text{电子产品} (0.7), \text{公司} (0.2), \text{电影} (0.08), \text{水果} (0.016) \dots \rangle$ , 其中各子意图概率之和为 1。

[0336] 3) 基于步骤 1) 和步骤 2) 的计算, 使用本体技术给出清晰、明确、可计算的用户情境说明和描述, 可以得知该用户的情境表示为 :

[0337]  $UC = \langle \{O_1, O_2\}, O_3, \langle \text{电子 产 品} (0.7), \text{公 司} (0.2), \text{电 影} (0.08), \text{水 果} (0.016) \dots \rangle, \{\text{苹果}\}, \{\text{吉林省长春市吉林大学}\}, \{2012-12-11\} \rangle$ 。

[0338] 3. 根据第 1 步骤中实现的网页情境表示设计融入网页情境信息的索引结构并保存到索引数据库中, 步骤如下 :

[0339] 1) 构建情境正向索引 :

[0340] 图 9 中所示的网页所对应的情境正向索引表示为 :

[0341] 2029- > {<Jefferies, Wall>, 0, 0.005, 0.112} - > {< 分析师, Db01>, 1, 0.02, 0.169} - > ..... - > {< 苹果, Dp03>, 6, 0.08, 0.721} - > .....

[0342] 其中 Wall、Db01 等代表词所对应的语境词在通用本体中的编号。

[0343] 2) 构建情境倒排索引

[0344] 利用情境正向索引,构造<词,语境词>与文档集合之间的索引,例如在第 1 步骤中实现的网页情境解析基础上,图 9 中所示的网页中的苹果一词可表示为:

[0345] < 苹果, Dp03> - > [2029] - > [125451] - > [184573] - > .....

[0346] 其中 Dp03 代表词“苹果”所对应的语境词在通用本体中的编号。

[0347] 为支持对时间和空间敏感的查询,需要借鉴时空数据库的索引结构对上述提出的索引进行进一步的改进,空间索引 GI 用 R\* 树构建,时间索引 TI 用 B+ 树构建,语境索引 CI 用倒排形式构建。

[0348] 将构建好的情境索引保存到索引数据库中。

[0349] 4. 由情境扩展模块实现情境扩展并形成查询情境,步骤如下:

[0350] 根据用户提供的查询词为“苹果”,结合在步骤 2 中用户情境挖掘与表示模块所获得的用户情境信息,使用通用本体提取查询词“苹果”的若干释义为  $\langle QS_1, QS_2, QS_3, QS_4, QS_5 \rangle$ ;

[0351]  $QS_1$ : 苹果,一种常见的水果。

[0352]  $QS_2$ : 苹果公司,一家著名的电子产品生产厂商,原名苹果电脑。

[0353]  $QS_3$ : 苹果日报,一份由香港上市公司壹传媒集团旗下的正体中文报纸。

[0354]  $QS_4$ : 苹果(电影),一部在 2007 年上映的中国电影。

[0355]  $QS_5$ : 苹果(产品),iphone、ipad 等。

[0356] 及其相应的若干语境  $\langle C_1, C_2, C_3, C_4, C_5 \rangle$  为:

[0357]  $C_1$ : 水果,  $C_2$ : 组织,  $C_3$ : 读物,  $C_4$ : 电影,  $C_5$ : 电子设备。

[0358] 2) 参阅图 8,基于步骤 2 中用户情境挖掘与表示模块得到的某用户情境信息,其长期兴趣为图中的  $O_1$  和  $O_2$ ,短期兴趣为图中的  $O_3$ ,当该用户提交查询词“苹果”时,利用朴素贝叶斯模型计算语境词  $C_i$  的后验概率  $p(C_i | I_1, I_2, \dots, I_n)$ ,经过计算,限定最佳语境词为“电子产品”

[0359] 3) 若用户选择某情境扩展,则取相应语境为最终确认结果,否则自动选取  $p(C_i | I_1, I_2, \dots, I_n)$  值最大的语境词“电子产品”为最终确认结果。用户选定扩展的语境词后,如果用户选定的语境词与本发明提供的最佳语境词不一致,需要更新该用户的兴趣和查询意图。

[0360] 4) 这样搜索结果页中就不再出现诸如“水果”、“电影”类网页。最后得到的查询情境为:

[0361] < {< 苹果, { 苹果 }, { 电子产品 }>, 吉长春市吉林大学, 2012-12-11}

[0362] 5. 由情境查询处理模块实现情境查询处理,步骤如下:

[0363] 1) 情境推荐:

[0364] 此步骤发生在用户打开查询界面但是没有进行查询操作时,对于经步骤 1 网页采集存储在网页数据库中网页,根据网页情境与用户情境的匹配程度及网页权威性,以合理的顺序向用户主动推荐网页,达到“不搜亦得”的智能化检索目标。情境推荐的形式化表示

为：

[0365] Result = Recommend (Pages, UT, UG, A, R<sub>SL</sub>)

[0366] 其中：Pages 表示新抓取的网页集合，UT 表示用户最近查询时间，UG 表示用户所处地点，A 表示网页权威性，R<sub>SL</sub> 表示用户兴趣与网页情境的相关性。

[0367] 参阅图 8，通过上述用户情境挖掘与表示模块中得到的某用户情境信息，其长期兴趣为 O<sub>1</sub> 和 O<sub>2</sub>，短期兴趣为 O<sub>3</sub>，所以根据网页情境与用户情境的匹配程度、用户兴趣与网页情境的相关性以及网页权威性，本发明将主动推荐电子产品类的网页和与当前的体育比赛相关信息的网页，以及有关美国总统奥巴马最近的新闻事件。

[0368] 对于该用户，本发明所给出的情境推荐明细如表 1 所示：

[0369] 表 1 情境推荐明细

[0370]

### 电子产品

苹果官网 - Apple 设计并创造了 iPod 和 iTunes、Mac 便携式和台式电脑、OS X 操作系统以及革命性的 iPhone 和 iPad.

【苹果专区】iPhone、iPad、Touch(报价\_图片\_降价)\_手机中国

Apple - iPhone 5 - 更薄,更轻,更快,更好的 iPhone

Apple (苹果) 官方客服电话

苹果不会提前派息：70%现金流来自海外

小米手机20万台开放购买 - 小米手机官网最新小米2及小米1s青春...

苹果-京东商城

苹果论坛-iPhone5论坛\_苹果5\_iPhone4\_iPhone4s\_iPhone5\_iOS6越狱\_...

### 体育

C罗卡卡领衔国王杯名单4将伤缺 朴智星再战皇马

卡西首次表态有意离皇马 佛爷弃他选定终身队长

美媒体：火箭凌厉林书豪表演 马刺尼尔得分新高

组图：哈登领5将服务小球迷 火箭能坐购物车搞怪

霍勒迪25+5+8特纳准三双 76人胜活塞后东部第6

NBA赛程直播中心

联赛杯-队长扳平+罚丢点球 阿森纳遭英乙队淘汰

曼城中场大将遭官方指控 后防屌丝逆袭夺月最佳

### 时事

前11月财政收支解读：收入开始回升 支出偏重民生

财政12月恐突击花钱2万亿 接近一季度支出总额

吉林省依托科技创新 力促经济平稳快速发展

奥巴马与众议院共和党领袖交换财政悬崖新提案

美国政府官员称总统奥巴马和共和党发言人举行电话交谈

中国新领导层向美传达维持稳定双边关系基调

马尔代夫副总统会见黎巴嫩新闻部长瓦利德 达欧格

美国伊利诺伊州一架医疗直升机坠毁 3人死亡

[0371] 2) 情境检索：

[0372] 此过程发生在用户提交了此次查询的查询词之后。

[0373] (1) 经步骤 4 情境扩展模块处理获得查询情境信息为： $\langle \langle \text{苹果}, \{\text{苹果}\}, \{\text{电子产品}\} \rangle, \text{吉林大学}, 2012-12-11 \rangle$ ，利用这些信息，在本发明构建的情境索引数据库中进行检索。

[0374] (2) 同时访问步骤 1 建立的网页数据库，得到相匹配的网页，形式化表示为：

[0375] Result = Search (Index, 苹果, 电子产品, 苹果, 吉长春市吉林大学, 2012-12-11)。

[0376] 用“电子产品”和“苹果”在索引 Index 中的第一层语境索引 CI 中进行检索，用用户所处地点“吉林省长春市吉林大学”在第二层空间索引 GI 中进行检索，用用户最近检索

时间“2012-12-11”在第三层索引时间 TI 中进行检索。

[0377] 3) 情境排序：

[0378] 此步骤发生在情境检索取得返回结果之后,对于该返回结果,依据情境相关性(语境相关性、地理位置、时间等)及网页权威性进行排序,排序之后的结果就可以提供到用户查询界面中了,情境排序的形式化表示为:

[0379]  $\text{Result}' = \text{Rank}(\text{Result}, A, R)$

[0380] 其中 :A 为网页权威性,采用经典的基于网页链接的 PageRank 静态排序算法,R 为情境相关性,是用户情境 UC 与网页情境 PC 匹配的结果,即 : $R = \text{Match}(UC, PC)$ 。该公式基于 Q 与 W 的匹配、I 与 C 的匹配、UG 与 PG 的匹配、UT 与 PT 匹配,返回相应网页的情境相关性值。

[0381] 例如:通过上面的分析,可以得知该用户对电子产品感兴趣,当用户的查询词为“苹果”时,那么通过 UC 和 PC 中的情境信息匹配情况,与电子产品相关的网页的排序位置会相应得到提升。

[0382] 如果该用户的兴趣是水果,则那么通过 UC 和 PC 中的情境信息匹配情况,并且考虑到该用户的 IP 地址是吉林省长春市,则查询返回的结果中应该将反映“最新吉林苹果价格”的网页信息(网页 2)显示在其他结果前面。

[0383] 网页 2

[0384]

最新吉林苹果价格 一亩田农产品商务平台			
批发市场价格			
产品名称	市场地区	12月11日均价	环比
苹果	山东日照苹果专业合作社	1.00元/斤	0%
苹果	吉林贸发	4.59元/kg	↓4.97%
苹果	吉林长春市	4.00元/kg	↓17.18%
苹果	吉林辽源市	3.20元/kg	↓33.75%
苹果	吉林白城市	3.20元/kg	↓33.75%

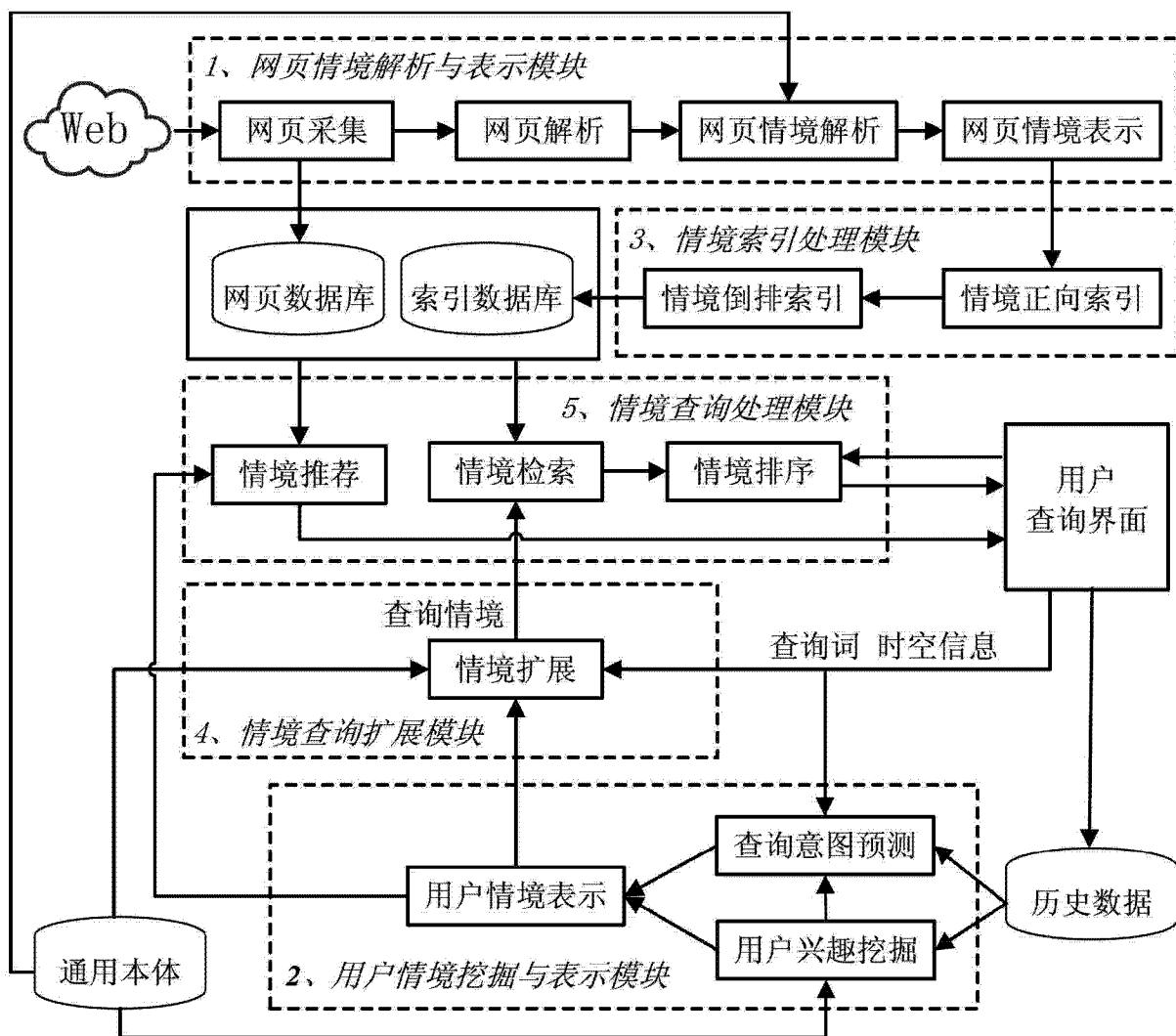


图 1

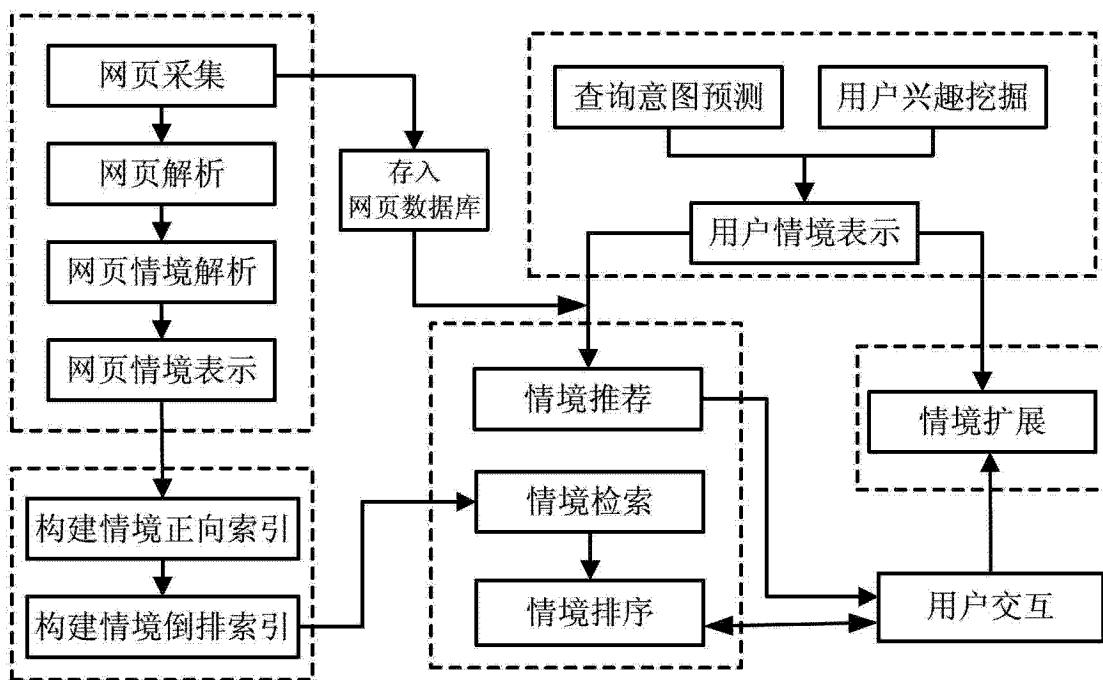


图 2

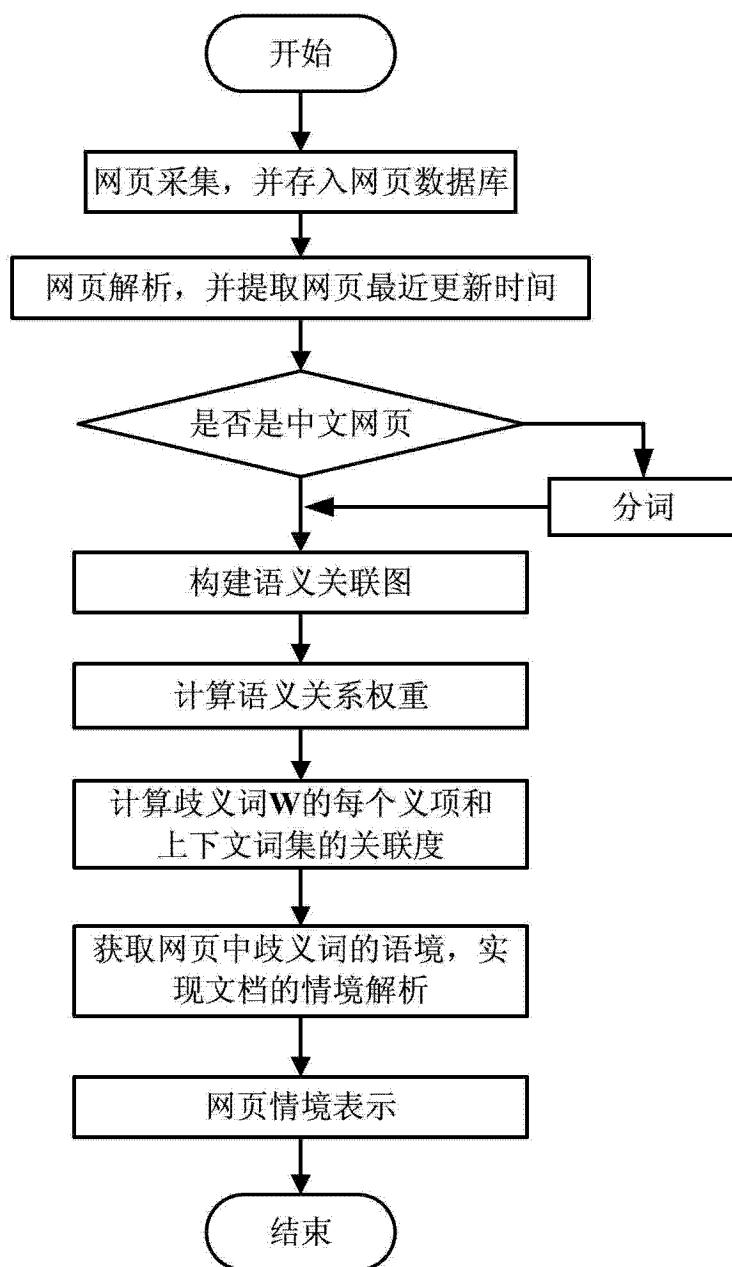


图 3

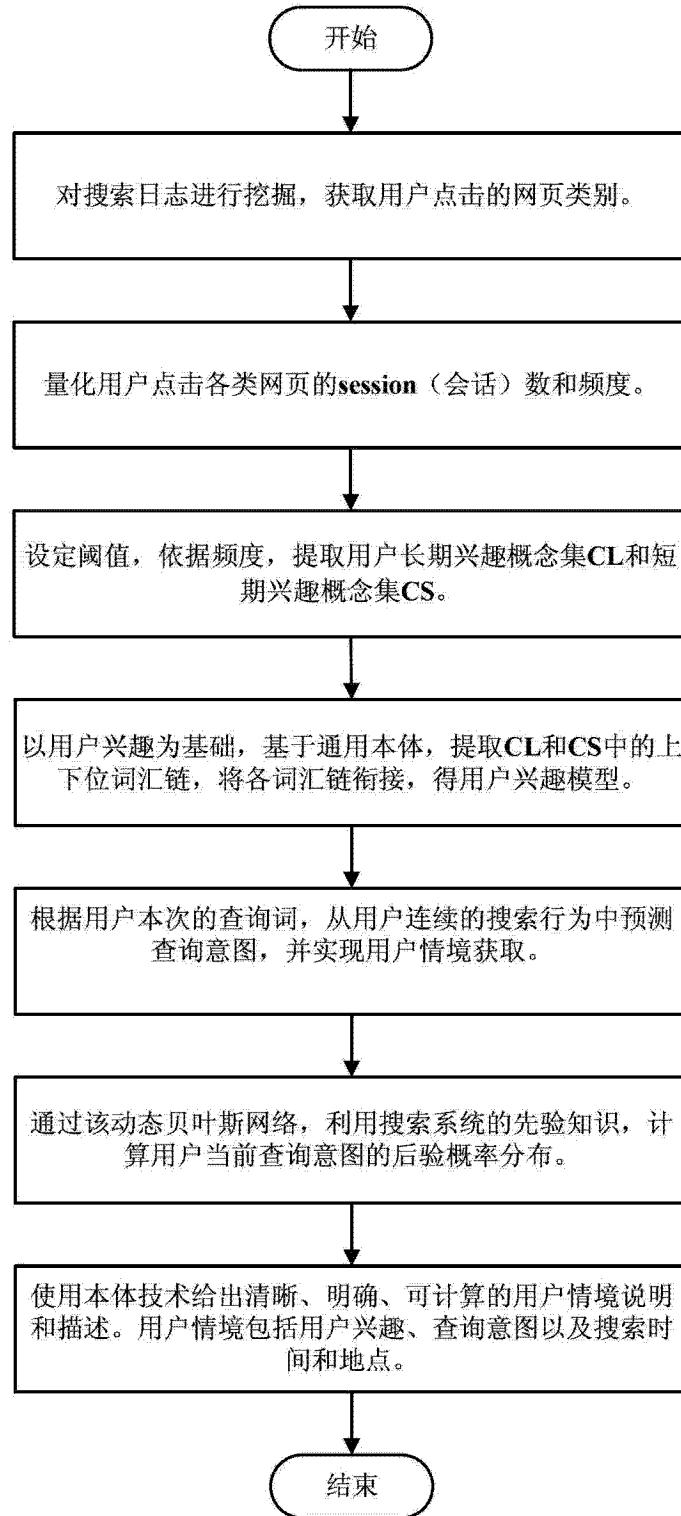


图 4

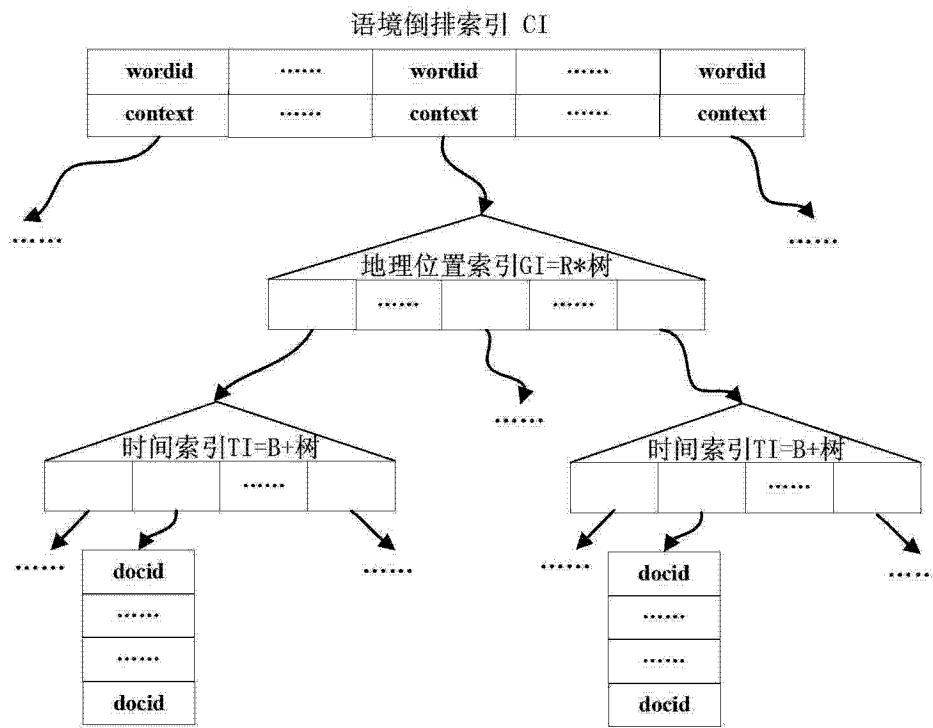


图 5

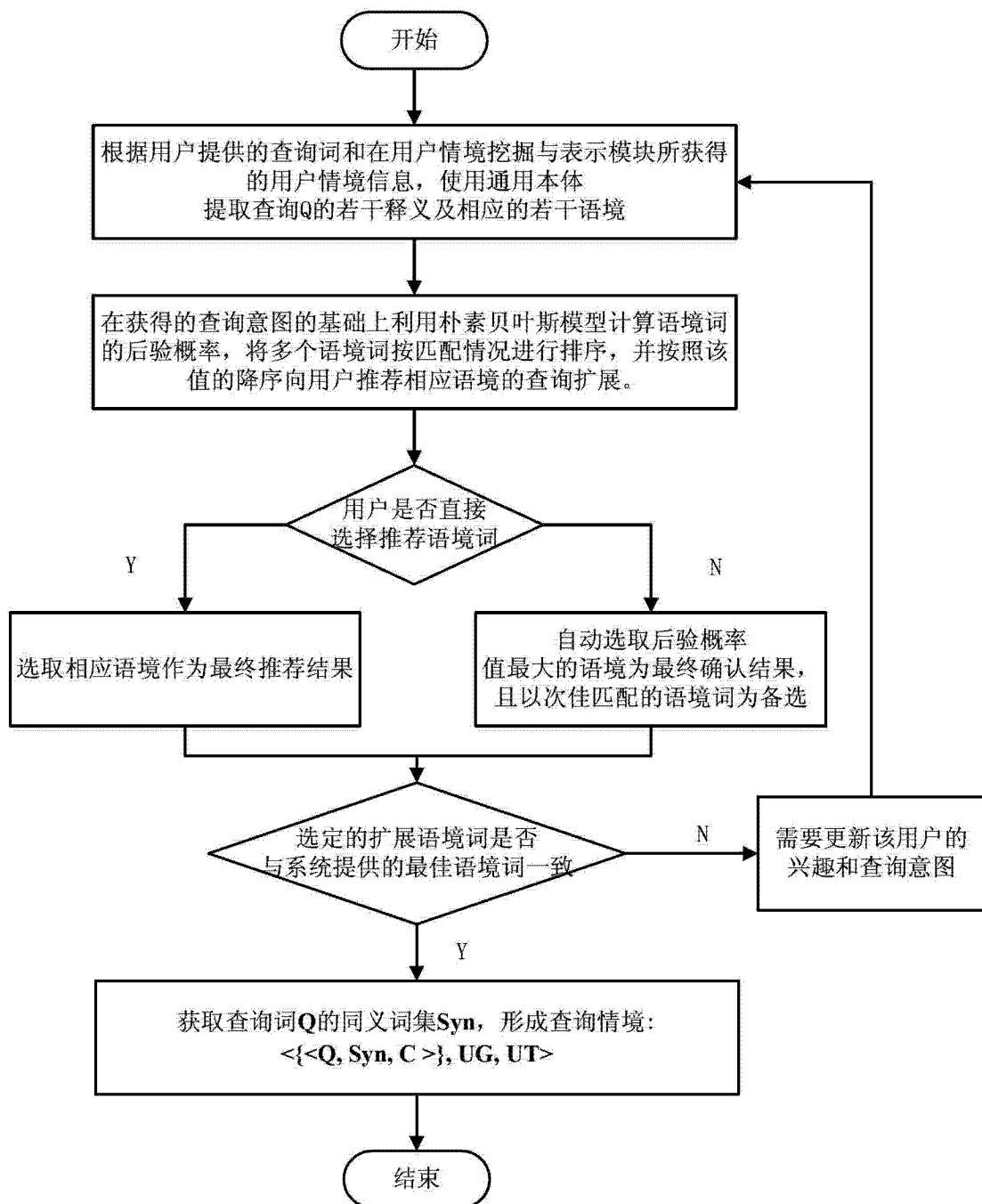


图 6

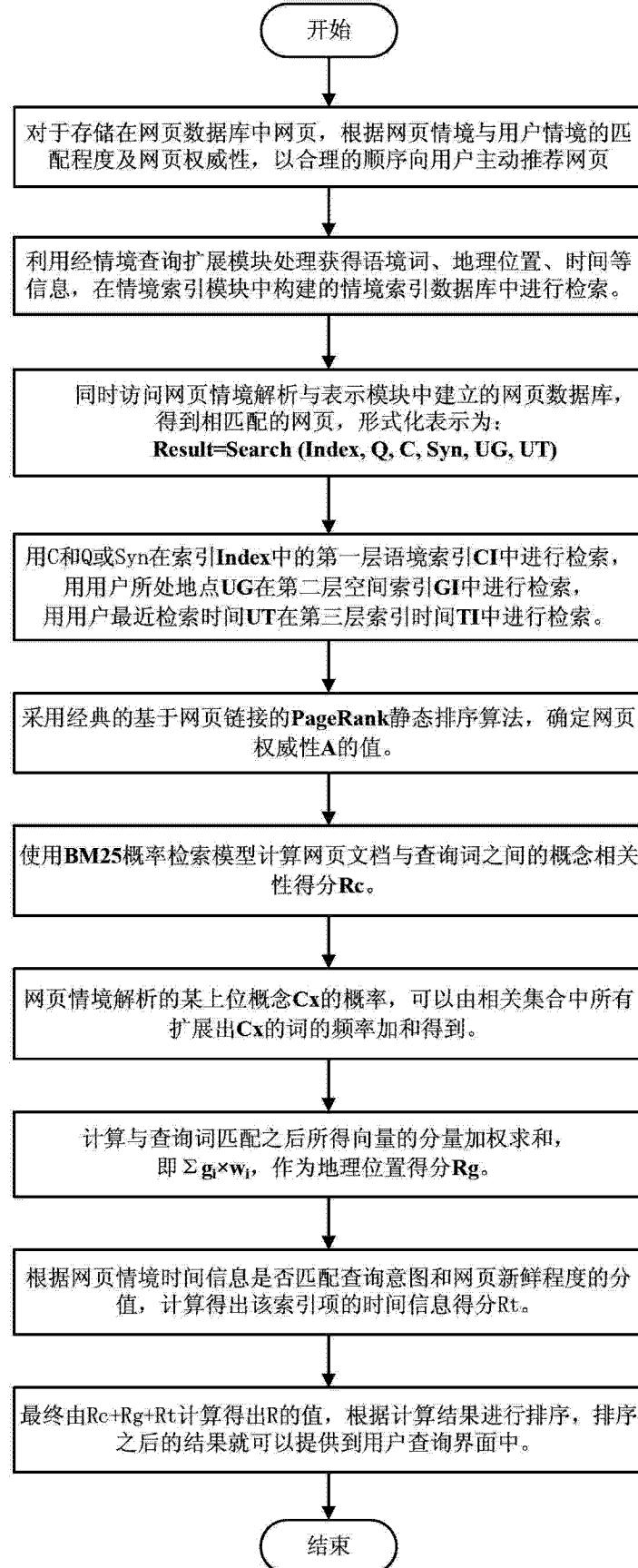


图 7

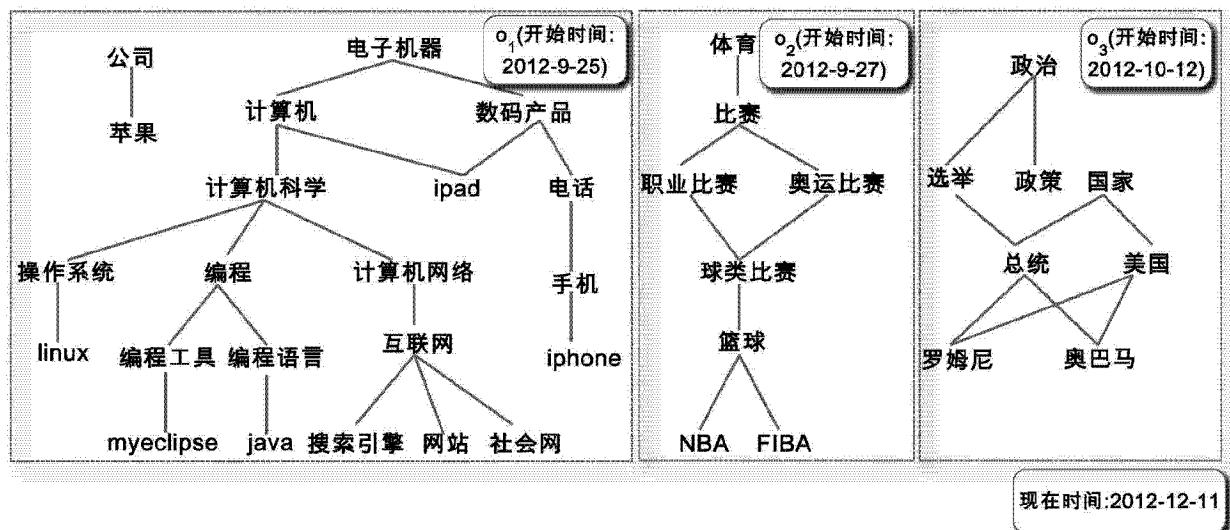


图 8

**sina 苹果汇**

新浪手机 | 新浪科技 | 新浪首页 | 新浪导航

首页 iPhone iPad Mac 公司新闻 应用游戏推荐 实用技巧 iOS商店下载

分析师称iPhone 5S六月推出：拥有多彩机身

2012年12月11日 08:37 新浪科技·苹果汇 作者：琪琪

中国大陆行货版iPad mini体验

相比9.7英寸的Pad, iPad mini重量轻了一半,但却能几乎完成和大号iPad完全一样的任务,且更便携.....[详细>>](#)

解析苹果“美国制造”战略

苹果之所以决定将部分电脑的生产带回美国国内,主要是顺应行业趋势.....[详细>>](#)

如果重新设计iOS系统解锁界面

苹果的iOS系统对锁屏界面的利用并不充分,尤其对于该平台数万app的整合来说功能很不到位.....[详细>>](#)

**苹果汇问答**

① @杰克79: 电话上的勿扰模式打开以后有什么作用。[@苹果汇](#)

② @苹果汇: 打开勿扰模式,可在指定时间内,收到电话、短信或者都不会点亮屏幕或是响铃,这对半夜“响一声电话”很有用。如果打开勿扰模式又怕漏接电话,它还提供了“置重来电”功能,比如某人在3分钟之内连打两个电话,第二个电话铃声仍会响起。也可将联系人添加到白名单,防止特别重要的电话漏掉。[查看详细>>](#)

图 9