

基于地理-时间意图和偏好的个性化 Web 搜索框架 GT-WSearch

杨 丹¹ 申德荣² 陈 默³

(辽宁科技大学软件学院 鞍山 114051)¹ (东北大学信息科学与工程学院 沈阳 110004)²

(东北大学计算中心 沈阳 110004)³

摘 要 基于 Web 查询的地理位置、时间查询意图和用户偏好的个性化 Web 搜索可以改善 Web 搜索结果,更好地满足不同用户的信息需求。提出了 GT-WSearch 个性化 Web 搜索框架,它通过挖掘搜索结果、用户点击数据和对查询进行分析得到的用户概貌和查询概貌,来捕捉用户的地理-时间的意图和偏好,提高搜索质量。用户概貌表明了查询自身的地理-时间的特性。GT-WSearch 框架在排序函数中利用文档的地理位置、时间的相关度来进行个性化搜索。最后将使用线性的相关度排序函数进行重新排序的搜索结果返回给用户。大量实验结果表明,所提出的个性化方法在提高 Web 搜索结果的质量中取得了明显的效果。

关键词 个性化 Web 搜索,地理-时间意图,用户偏好

中图分类号 TP311.13 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.7.051

Geo-temporal Intent and Preference-based Personalized Web Search Framework GT-WSearch

YANG Dan¹ SHEN De-rong² CHEN Mo³

(College of Software, University of Science and Technology Liaoning, Anshan 114051, China)¹

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)²

(Computing Center, Northeastern University, Shenyang 110004, China)³

Abstract Traditional Web search ignores user's geographic (geo) and temporal preference. Personalized Web search results according to query's implicit geo-temporal intent and preferences can help to satisfy users' different information needs greatly. GT-WSearch framework was proposed to capture users' geo-temporal preferences leveraging both query profile and user profile. The query profile signals the geo-temporal characteristic of the query itself. The user profile is obtained by mining not only search results snippets but also user click-through data. GT-WSearch considers geo-temporal relevance of documents when re-ranking search results. The experimental results show that our personalizing approach effectively improves the quality of Web search results.

Keywords Personalized Web search, Geo-temporal intent, User preference

随着 Web 数据,尤其是在线社交网络数据的不断增大和更新,大多数的 Web 数据都包含显式或隐式的地理位置、时间信息元数据。例如,社交媒体信息中,大多数都包含文字或 GPS 报告的位置信息,同时这些 Web 文档至少有创建时间作为它的时间上下文。因此如何根据用户潜在的地理位置、时间意图和偏好个性化 Web 搜索结果,成为 Web 搜索需要解决的热点问题之一。多数时候当用户发出一个 Web 查询时,通常在他/她脑中可能有一个地理位置或时间上的意图和偏好。以前的研究表明,13%以上的 Web 查询包含地理信息^[1-3],并且 83.77%是包含城市信息的明确的地理查询(geo query)^[3]。此外,文献[4]分析显示,超过 7%的查询都属于隐式的年查询(year qualified query)。这说明用户提交的 Web 查询中,带有地理位置、时间意图和偏好的比例占了很大的部分。但是当前大多数的商业 Web 搜索引擎对所有的用户返

回相同的结果,并没有考虑用户的地理位置、时间意图和偏好。下面通过几个 Web 搜索的例子进一步说明基于地理位置、时间意图和偏好的个性化 Web 搜索需要解决的问题。

例 1 用户提交查询“Beijing, Job”想要查找地点在北京的工作信息,而不是查找关于介绍“北京”的 Web 文档。类似地,用户输入查询“Hotel, shanghai”想要搜索在上海的宾馆的信息,而不是搜索包含“shanghai”关键字的文档。

例 2 用户输入查询“weather forecast”查找他/她所在城市如沈阳的天气预报信息。类似地,用户输入查询“SIGMOD”想找的很可能是即将召开的 2014 年 SIGMOD 会议的相关信息,或是以前某一年的 SIGMOD 会议信息。

例 3 一个北京的用户在圣诞节期间输入 Web 查询“department, discount”想要查找北京的商场将来某个时间(段)的打折信息,而不是暑假期间的打折信息。

到稿日期:2014-07-07 返修日期:2014-11-27 本文受国家自然科学基金项目(61402213,61402093),中央高校基本科研业务费专项资金项目(N120316001),辽宁省教育厅科学研究项目(L2013120)资助。

杨 丹(1978-),女,博士,副教授,CCF 会员,主要研究方向为数据集成、数据空间,E-mail:asyangdan@163.com;申德荣(1966-),女,教授,博士生导师,CCF 高级会员,主要研究方向为分布式数据库;陈 默(1983-),女,博士,讲师,CCF 会员,主要研究方向为空间数据处理。

通过上述例子可知, 大多数时候用户想搜索的是某个特定地理位置的对象, 如饭店、工作信息、购物中心等, 或特定时间的事件, 如某年的世界杯、某年的数据库会议等。不同用户具有不同的地理位置、时间上的意图和偏好。因此需要的解决的问题主要包括: (1) Web 用户可以通过在查询中使用地名、时间来显式地表示他们在地理位置、时间上的查询意图, 但是对于缺少显式地理位置、时间信息的 Web 查询, 如何捕捉用户潜在的地理位置、时间意图和偏好进行个性化? (2) 对于有显式地理位置、时间信息的 Web 查询, 如何将关键字中的地理位置、时间信息与内容部分进行区分? (3) 如何在搜索结果的个性化排序中考虑地理位置、时间的相关度, 提高用户的信息满意度来更好地个性化 Web 搜索?

本文提出基于地理-时间意图和偏好的个性化 Web 搜索框架 GT-WSearch, 通过分析和挖掘 Web 查询自身、Web 查询的搜索结果和用户点击数据, 既考虑用户概貌 (user profile) 同时又考虑查询概貌 (query profile) 来捕捉用户的地理-时间意图和偏好进行 Web 搜索结果个性化。

1 相关工作

本节首先介绍与本文研究相关的地理查询 (geo query)、时间查询 (temporal query) 及个性化的相关工作; 接着介绍地理-时间搜索 (Geo-temporal search); 最后比较了 GT-WSearch 与上述相关工作的不同。

文献[5-8]的相关工作是 Web 查询的地理意图发现或根据地理位置进行 Web 搜索结果个性化。文献[5]内置一个城市的语言模型, 从大量的网页搜索记录中发现、学习并建立一个地理意图的分析系统。文献[6]通过挖掘搜索结果和用户的点击数据来捕捉用户的地理位置偏好进行个性化 Web 搜索。文献[7]建立一个概率模型来识别潜在的位置意图的查询, 利用用户的地理位置来提高 Web 搜索结果的相关度。文献[8]将社交媒体数据的空间搜索 (spatial search) 分为具有显式位置词的空间关键字搜索 (spatial keyword search) 和隐式的位置意图搜索 (local intent search)。

在信息检索中, 时间查询^[9]通常是为了找到最新的结果或实体的世系信息。文献[4, 10-12]考虑了 Web 查询的时间维度。文献[4]提出一种从查询日志中挖掘隐式年份的查询 (year qualified queries) 的方法。文献[10]提出识别隐式时间的查询并重新排序搜索结果。文献[11]提出基于时间信息的文档摘要来支持基于时间的搜索。文献[12]提出了 3 种类型的时间查询, 即时间的 (temporal)、时间明确的 (temporally unambiguous) 和时间模糊的 (temporally ambiguous) 查询。文献[13]根据查询之前在日志里出现的情况将查询分类成具有隐式时间的意图和具有显式时间表达式的查询。文献[14]讨论了结果的时间聚类。文献[15]出现了使用时间的语言模型来决定查询的时间的几种方法, 利用决定的查询的时间进行搜索结果重排序来提高信息搜索的有效性。文献[16]基于查询随时间的流行度的变化将它们分类, 并为查询流行度的变化和 Web 内容的更新提供了检测查询意图变化的有用信号。而目前并没有支持在时间轴上的关键字搜索的搜索系统。

地理-时间的搜索 (Geo-temporal search) 是指那些具有地理的和时间的限制的搜索。尤其是对事件或回答关于特定地

点、时间事件的问题的搜索^[17]。近几年随着社交媒体的普及, 与地理-时间的搜索相似, 出现了社交媒体数据的空间-时间的 (Spatio-temporal) 搜索^[8, 18, 19]或分析。文献[8]给出 3 种空间-时间查询的类型, 但是并没有给出具体的实现方法。文献[18]建立了叫做 MoboQ 的应用, 用户可以提出时间的、地理敏感的问题, 然后及时地从其他用户那里得到众包 (crowdsourced) 的答案。但是通过前文的例子分析可知, 地理-时间的搜索普遍存在于 Web 搜索中, 并不仅限于对事件的搜索, 也不仅限于社交媒体数据中。在文档重新排序方面, 文献[20]结合文本的、时间的和地理的查询给出了一种新的文档排序模型, 除了计算查询中的文本、时间和地理各个部分的相关度得分, 还计算了接近度得分 (proximity score), 从而消除了这 3 部分之间是独立的这种简单假设带来的问题。

已有的相关工作中时间的信息检索、地理的信息检索经常被分开考虑, 要么只考虑地理的查询, 要么只考虑时间的查询, 而 GT-WSearch 考虑了与用户 Web 查询相关联的地理位置、时间的信息需求和偏好。此外, 不同于大多数的个性化搜索方法, GT-WSearch 不仅考虑了用户概貌, 还考虑了 Web 查询自身的地理、时间的特征 (在本文中被称为查询概貌) 来进行个性化 Web 搜索。

2 GT-WSearch 概览

本文提出的基于地理位置、时间的个性化 Web 搜索框架 GT-WSearch 如图 1 所示。GT-WSearch 由查询处理、查询概貌管理、用户概貌管理和个性化排序模块组成。对于用户输入的一个 Web 查询 Q , 首先利用查询处理模块 (Query Processing) 对其查询意图进行判定; 接着提交该查询到后端的搜索引擎 (已有的商业搜索引擎) 得到例如 TOP 50 的搜索结果 R ; 通过查询概貌管理模块 (query profile management) 和用户概貌管理模块 (user profile management) 分别生成查询概貌和用户概貌; 最后基于查询概貌和用户概貌通过个性化模块对搜索结果 R 进行重排序后将个性化的 Web 搜索结果 R^+ 返回给用户。

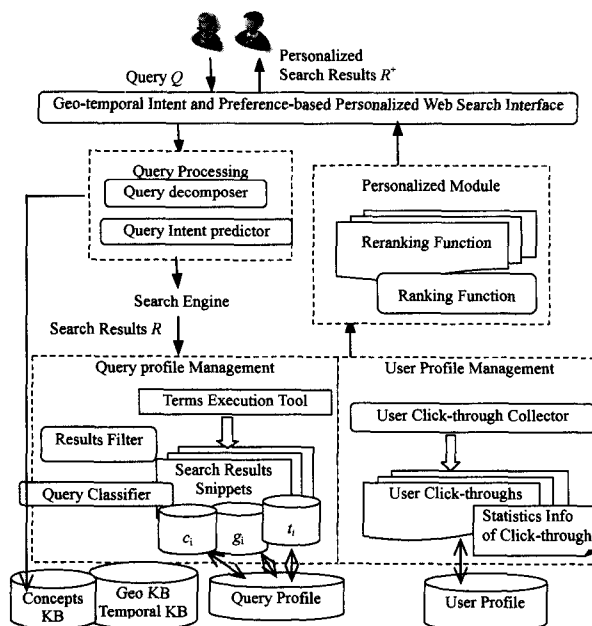


图 1 个性化 Web 搜索框架 GT-WSearch

GT-WSearch 引入用户概貌和查询概貌来捕捉用户的地理位置和时间偏好进行个性化查询。前者是通过挖掘和分析用户点击的数据来获得用户地理和时间偏好,而后者是通过挖掘 Web 查询的搜索结果和 Web 查询自身而得到 Web 查询内涵。

3 基于地理位置和时间的个性化 Web 搜索

3.1 查询处理

不同的 Web 查询本身具有不同的地理位置、时间特性。例如,查询“renting house”、“abroad travel”对相关地理信息的要求要大于时间信息。而查询“newest movies introduction”是一个既具有地理信息要求又具有时间要求的查询。但查询“java programming”对内容的要求比地理和时间更高。查询概貌存储 Web 查询自身的内涵信息,表示一个查询自身与生俱来的对内容、地理位置和时间的信息要求和包含能力。

查询意图判定的过程包括 3 个步骤。步骤 1:匹配、映射,利用系统底层的知识库(即概念知识库(Concept KB)、地理位置知识库(Geo KB)和时间知识库(Temporal KB),将查询 $Q = \{k_1, k_2, \dots, k_{|Q|}\}$ 匹配和映射到内容(记作 Q_C)、地理位置(记作 Q_G)和时间(记作 Q_T)3 个部分,即 $Q = \{Q_C, Q_G, Q_T\}$ 。匹配和映射的结果可能缺少地理部分或(和)时间部分,即 $Q_C = \text{null}, Q_T = \text{null}$ 。 Q_C 是查询 Q 的内容特征, Q_G 是查询 Q 的地理位置特征, Q_T 是查询 Q 的时间特征。步骤 2:转换、标准化,将地理部分和时间部分进行格式转换、标准化等。步骤 3:查询意图解释,根据 Q_C 和底层的概念知识库对查询 Q 的潜在查询意图进行解释。

3.2 查询概貌管理

查询概貌的生成过程主要由 2 个步骤组成。步骤 1:从 Web 搜索结果摘要中提取查询的特征词,即内容、地理位置和时间词汇;步骤 2:使用查询分类器对查询进行分类。

3.2.1 搜索结果分析

从 Web 搜索结果摘要(Snippet)中提取内容词/短语(记为 c_i)、地理词(记为 g_i)和时间词/表达式(记为 t_i)。采用式(1)提取每个查询 Q 的 Top 50 搜索结果摘要中的所有内容词 c_i 。

$$F(c_i) = \frac{f(c_i)}{n} |c_i| \quad (1)$$

其中, f 是词 c_i 的频率(即含有 c_i 的文件数目), n 是返回的文档的数量, $|c_i|$ 是 c_i 中单词(term)的个数。如果一个词 c_i 的 F 值高于阈值 θ (实验中设置 $\theta=0.04$),则把 c_i 作为查询 Q 的一个内容词。

考虑到位置词较稀疏,位置词提取工具利用框架底层的地理知识库提取全部搜索结果文档(而不是搜索结果摘要)中的所有地理位置词。地理位置词主要包括:地名(城市、地区和国家等)和地理特征(岛、河、湖和海等)。目前 GT-WSearch 框架没用考虑方向和空间介词(如“South”、“nearby”和“between”等)。类似地,考虑到时间词稀疏,时间词提取工具利用框架底层的时间知识库提取全部搜索结果文档中的所有时间词(表达式),如“May”、“Christmas”、“2014”等。由于文档中时间特征词汇的复杂性,目前 GT-WSearch 侧重于以下时间粒度,即年、月、日,没有考虑相对含义的时间词汇,如“today”、“tomorrow”、“recently”等。

3.2.2 查询分类

根据提取的 Web 查询 Q 自身的内容、地理位置、时间词的结果(这里被称作 Web 查询内涵),查询分类器使用决策树算法将 Web 查询分类。引入地理位置信息熵率(记作 R_g , 见式(2))和时间熵率(记作 R_t , 见式(3))的概念,它们分别是 Web 查询的地理位置信息熵、时间信息熵与内容信息熵的比率。其中,离散随机变量 X 的信息熵的定义见式(4), $X = \{x_1, x_2, \dots, x_n\}, p(x_i) = pr(X=x_i)$ 。

$$R_g = \text{entropy}(g_i) / \text{entropy}(c_i) \quad (2)$$

$$R_t = \text{entropy}(t_i) / \text{entropy}(c_i) \quad (3)$$

$$\text{entropy}(X) = - \sum_{i=1}^n P(x_i) \log p(x_i) \quad (4)$$

接着根据 R_g 和 R_t 使用决策树算法将 Web 查询分为 5 类(记为 Q1-Q5),图 2 给出了一个用于查询分类的决策树模型例子。其中,Q1 类查询与地理位置和时间关联都较高,如查询“department, discount”;Q2 类查询与地理位置信息关联较高,但与时间信息关联较低,如查询“Weather forecast”;Q3 类查询与位置信息关联较低,但与时间信息关联较高,如查询“stock, price”、“SIGMOD”;Q4 类查询对地理位置信息和时间信息的要求都相对较低,如查询“Dell, notebook”;Q5 类查询对地理位置信息和时间信息都不敏感,如查询“java programming”、人名、一些专有名词等。

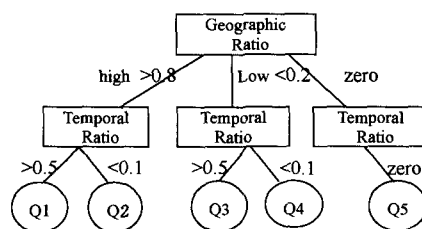


图 2 Web 查询分类的决策树模型例子

3.3 用户概貌管理

即使同一个 Web 查询,不同的用户也具有不同的内容、地理位置和时间信息要求。通常,用户对搜索结果的点击动作反映出他/她的查询意图和偏好。例如,用户提交 Web 查询“lotus, 2014, beijing”,其中关键字“lotus”的概念方面组成包括软件、汽车、植物和酒等。如果用户的意图是搜索北京 4S 店的 2014 年莲花牌汽车的信息,那么他/她可能会点击具有内容“汽车”或“跑车”等、地理位置“北京”和时间“2014”这些词汇的相应搜索结果。因此,GT-WSearch 从用户点击的统计结果中提取的内容、地理位置和时间词(表达式)作为用户的内容、地理位置和时间偏好空间。利用它们的点击熵(见式(5)一式(7))作为用户的查询意图和偏好系数。

$$PRE_c = - \sum_{i=1}^n \frac{|c_i|}{|C|} \log \frac{|c_i|}{|C|} \quad (5)$$

$$PRE_g = - \sum_{i=1}^m \frac{|g_m|}{|G|} \log \frac{|g_m|}{|G|} \quad (6)$$

$$PRE_t = - \sum_{i=1}^k \frac{|t_k|}{|T|} \log \frac{|t_k|}{|T|} \quad (7)$$

其中, n 是用户点击内容词的个数, $C = \{c_1, c_2, \dots, c_n\}, |c_i|$ 是用户点击内容词 c_i 的次数, $|C| = |c_1| + |c_2| + \dots + |c_n|$; m 是用户点击的地理位置词的个数, $G = \{g_1, g_2, \dots, g_m\}, |g_m|$ 是用户点击地理位置词 g_i 的次数, $|G| = |g_1| + |g_2| + \dots + |g_m|$; k 是用户点击的时间词的个数 $T = \{t_1, t_2, \dots, t_k\}, |t_i|$ 是

用户点击时间词 t_i 的次数, $|T| = |t_1| + |t_2| + \dots + |t_k|$ 。

3.4 基于地理位置和时间的个性化排序

使用加权求和的线性模型重新定义文档 d 对于查询 Q 的排序分值 $S(Q, d)$ (见式(8))。式中前3项是假设查询的内容、地理位置和时间3部分是相互独立的情况下, 它们分别的排序分值。而有些查询事件的地理位置、时间与内容是相互依赖的, 如查询“Olympic, 2008, Beijing”。因此对于此类查询, 在相关度得分中减去这部分的得分(最后一项)。

$$S(Q, d) = \alpha PRE_c S(Q_c, d) + \beta PRE_g S(Q_g, d) + \lambda PRE_t S(Q_t, d) - \frac{\beta + \lambda}{2} S(Q_c, d | Q_g, Q_t) \quad (8)$$

式中, $S(Q_c, d)$ 是内容排序分值, PRE_c 是相关联的权重值, $S(Q_g, d)$ 是地理位置排序分值, PRE_g 是相关联的权重值, $S(Q_t, d)$ 是时间相关度, PRE_t 是相关联的权重值; $S(Q_c, d | Q_g, Q_t)$ 是地理位置、时间与内容相互依赖的查询的地理位置、时间排序分值之和; α, β 和 λ 分别是查询的内容、地理位置和时间部分的权重系数, 并且 $\alpha + \beta + \lambda = 1$ 。

4 实验

实验查询集包含8大类主题的100个查询, 涉及Q1-Q5类查询。这些查询具有显式或隐式的时间意图; 或者显式或隐式的地理意图。表1给出了实验查询集的统计情况。由实验室选出10人作为用户, 其中每个用户负责提交10个查询。用户点击数据集的统计信息如表2所列。

表1 实验查询集的主题

Category	Description	Category	Description
1	Academic conference	5	Computer science & Software
2	Famous people	6	Movie
3	History event	7	Sports
4	Traveling	8	Shopping

表2 用户点击数据集的统计情况

Number of users	Number of test queries	Number of clicks	Avg. clicks per query
10	100	3750	3.75

4.1 实验参数设置

根据查询概貌的不同类查询, 式(8)中 α, β 和 λ 分别设置不同的值, 如表3所列。直觉和经验告诉我们无论是哪类查询, 其内容的相关度都要比地理、时间的相关度重要, 因此设置 α 的值大于 β 和 λ 的值。

表3 α, β 和 λ 的设置

Query Class	Q1	Q2	Q3	Q4	Q5
α	0.6	0.6	0.6	0.8	1
β	0.2	0.3	0.1	0.1	0
λ	0.2	0.1	0.3	0.1	0

4.2 实验评价

搜索结果的排序质量: 实验比较了本文提出的个性化方法与两种 baseline 方法, 结果如图3所示。

- baseline1: 不使用个性化的方法, 即传统的 Web 搜索引擎返回的搜索结果;

- baseline2: 不考虑查询概貌的个性化方法。

从图3可以看出, 与两种 baseline 方法相比, GT-WSearch 查询平均准确率都有所提高, 尤其对于 Q1、Q2 和

Q3 类查询。同时可以看出, 不考虑查询概貌的方法虽然平均准确率要高于 baseline1 方法, 但是不如 GT-WSearch 方法, 即既考虑用户概貌也考虑查询概貌的方法, 这说明查询概貌对个性化具有影响。

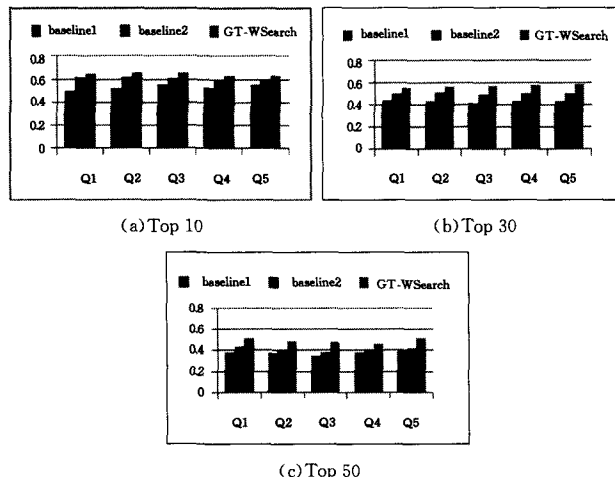


图3 Top 10, 30, 50 搜索结果的准确率比较

直观地说, 一个好的排序函数 $S(Q, d)$ 应该将搜索结果中用户想要的链接即用户想要的查询结果排序靠前。因此, 搜索结果中用户的平均点击越少说明其是质量更好的排序, 记作 $Avg. c(Q)$ 。因此, 较小的 $Avg. c(Q)$ 表示质量更好的排序。实验比较了个性化过程前后的 $Avg. c(Q)$, 结果如表4所列, 个性化后平均点击提高了 20.2%, 尤其 Q1、Q2 和 Q3 类查询的提高率较大。

表4 平均点击排序比较

Query Class	Avg. c(Q) (before personalized)	Avg. c(Q) (after personalized)	Improvement
Q1	3.8	2.9	24%
Q2	3.7	2.9	22%
Q3	3.7	2.8	24%
Q4	3.5	2.9	17%
Q5	3.5	3	14%

结束语 本文提出了一种结合用户概貌和查询概貌(即 Web 查询自身地理、时间特性), 通过挖掘 Web 搜索结果和用户的点击数据, 捕捉用户地理-时间的意图和偏好来进行个性化 Web 搜索的方法。实验结果表明, 所提出的方法可以提供给用户基于地理位置、时间意图和偏好的更个性化、更相关的 Web 搜索结果。

参考文献

- [1] Sanderson M, Kohler J. Analyzing geographic queries [C]// Proc. of ACM GIR. Sheffield, UK, 2004
- [2] Jones R, Zhang W V, Rey B, et al. Geographic intention and modification in web search[J]. International Journal of Geographical Information Science(IJGIS), 2008, 22(3): 229-246
- [3] Welch M J, Cho J. Automatically identifying localizable queries [C]// Proc of ACM SIGIR. 2008: 507-514
- [4] Metzler D, Jones R, Peng Fu-chun, et al. Improving search relevance for implicitly temporal queries[C]// Proc. of SIGIR. 2009: 700-701
- [5] Xing Yi, Raghavan H, Leggetter C. Discovering users' specific geo intention in Web search[C]// Proc. of WWW'2009. 2009:

- [6] Leung Wai-ting, Lee D L, Lee W-C. Personalized Web search with location preferences[C]//Proc. of ICDE'2010. 2010; 701-712
- [7] Lu Yu-mao, Peng Fu-chun, Wei Xing, et al. Personalize Web search results with user's location[C]//Proc. of SIGIR'2010. 2010;763-764
- [8] Leon R A D, Yang Bin, Christian S J. Towards context-aware search and analysis on social media data[C]//Proc. of EDBT 2013. 2013;137-142
- [9] Alonso O, Strogen J, Baeza-Yates R, et al. Temporal information retrieval; challenges and opportunities[C]//Proc. of TempWeb workshop. 2011;1-8
- [10] Zhang R, Chang Y, Zheng Z, et al. Search result re-ranking by feedback control adjustment for time-sensitive query[C]//Proc. of NAACL. 2009;165-168
- [11] Alonso O, Baeza-Yates R, Gertz M. Effectiveness of temporal snippets[C]//Proc. of WWW'. 2009
- [12] Jones R, Diaz F. Temporal profiles of queries [J]. ACM Transactions on Information System, 2007, 25(3)
- [13] Metzler D, Jones R, Peng F, et al. Improving search relevance for implicitly temporal queries[C]//Proc. of SIGIR. 2009;700-701
- [14] Alonso O, Strotgen J, Baeza-Yates R, et al. Temporal information retrieval; challenges and opportunities [C] // Proc. of TAWW. 2011;1-8
- [15] Kanhabua N, Novag K. Determining time of queries for re-ranking search results[C]//Proc. of European Conference on Digital Libraries(ECDL). 2010;261-272
- [16] Kulkarni A, Teevan J, Svore K M, et al. Understanding temporal query dynamics[C]//Proc. of WSDM. 2011;167-176
- [17] Gey F C, Kando N, Larson R R. The crucial Role of semantic discovery and markup in Geo-temporal search [C] // Proc. of CIKM. 2010;5-6
- [18] Kamath K, Caverlee J, Lee K, et al. Spatio-Temporal Dynamics of Global Social Media[C]//Proc. of WWW. 2013
- [19] Liu Ye-feng, Alexandrova T, Nakajima T. Using Stranger as Sensors; Temporal and Geo-sensitive Q&A via Social Media[C]//Proc. of WWW'. 2013
- [20] Strotgen J, Gertz M. Proximity2-aware ranking for textual, temporal, and geographic queries[C]//Proc. of CIKM. 2013
-
- (上接第 221 页)
- [43] Murdoch S J, Lewis S. Embedding covert channels into TCP/IP [C] // Information Hiding. Springer Berlin Heidelberg, 2005; 247-261
- [44] Mazurczyk W, Szczypiorski K. Evaluation of steganographic methods for oversized IP packets[J]. Telecommunication Systems, 2012, 49(2); 207-217
- [45] Sadeghi A R, Schulz S, Varadharajan V. The Silence of the LANs; Efficient Leakage Resilience for IPsec VPNs[M]//Computer Security-ESORICS 2012. Springer Berlin Heidelberg, 2012;253-270
- [46] Ji L, Liang H, Song Y, et al. A normal-traffic network covert channel [C] // 2009 Computational Intelligence and Security (CIS'09). IEEE, 2009; 499-503
- [47] Wendzel S, Keller J. Systematic engineering of control protocols for covert channels[C]//Communications and Multimedia Security. Springer Berlin Heidelberg, 2012; 131-144
- [48] Postel J. RFC 792; Internet control message protocol[Z]. 1981
- [49] Qu H, Su P, Feng D. A typical noisy covert channel in the IP protocol[C]//38th Annual 2004 International Carnahan Conference on Security Technology. IEEE, 2004; 189-192
- [50] Lucena N B, Lewandowski G, Chapin S J. Covert channels in IPv6 [C] // Privacy Enhancing Technologies. Springer Berlin Heidelberg, 2006; 147-166
- [51] Zander S, Armitage G, Branch P. An empirical evaluation of IP Time To Live covert channels[C]//15th IEEE International Conference on Networks(ICON 2007). IEEE, 2007; 42-47
- [52] Krätzer C, Dittmann J, Lang A, et al. WLAN steganography; a first practical review[C]//Proceedings of the 8th Workshop on Multimedia and Security. ACM, 2006; 17-22
- [53] 华元彬, 蒋建春, 卿斯汉. 基于链路分析法的复合隐蔽通道检测 [J]. 计算机应用, 2006, 26(1); 81-83
- Hua Yuan-bin, Jiang Jian-chun, Qing Si-han. Complex covert channel detection based on chain analysis methodology[J]. Computer Applications, 2006, 26(1); 81-83
- [54] 薛晋康, 许士博. 基于流量分析的网络隐蔽通道检测模型[J]. 计算机工程, 2002, 28(12); 46-48
- Xue Jin-kang, Xu Shi-bo. A network covert channel detecting model based on traffic analysis [J]. Computer Engineering, 2002, 28(12); 46-48
- [55] Yuwen Q, Huaju S, Chao S, et al. Network covert channel detection with cluster based on hierarchy and density[J]. Procedia Engineering, 2012, 29; 4175-4180
- [56] Sohn T, Seo J T, Moon J. A study on the covert channel detection of TCP/IP header using support vector machine[M]//Information and Communications Security. Springer Berlin Heidelberg, 2003; 313-324
- [57] Tumoian E, Anikeev M. Detecting NUSHU covert channels using neural networks[EB/OL]. http://www.ouah.org/neural_networks_vs_NUSHU.pdf, 2005
- [58] Tumoian E, Anikeev M. Network based detection of passive covert channels in TCP/IP[C]//The 30th Anniversary IEEE Conference on Local Computer Networks, 2005. IEEE, 2005; 802-809
- [59] Borders K, Prakash A. Web tap: detecting covert web traffic[C] // Proceedings of the 11th ACM conference on Computer and communications security. ACM, 2004; 110-120
- [60] 章思宇, 邹福泰, 王鲁华, 等. 基于 DNS 的隐蔽通道流量检测 [J]. 通信学报, 2013, 34(5); 143-151
- Zhang Si-yu, Zou Fu-tai, Wang Lu-hua, et al. Detecting DNS-based covert channel on live traffic[J]. Journal on Communications, 2013, 34(5); 143-151
- [61] Team S. Snort Users Manual 2.9.5[Z]. 2013
- [62] Cisco Systems, Inc. User Guide for Cisco Security MARS Local and Global Controllers, Release 6. x[Z]. 2014
- [63] Marleau G, Hebert A, Roy R. A User Guide for DRAGON Version 5[Z]. 2014