

УДК 004.89

Е.А. Дружинин, И.В. Шостак, А.А. Лысенко

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

АНАЛИЗ ПРОБЛЕМЫ ПОДДЕРЖКИ ФУНКЦИОНИРОВАНИЯ УНИВЕРСИТЕТСКИХ КАМПУСНЫХ СЕТЕЙ

Приведено определение университетской кампусной сети с определением её назначения и решаемых задач. Рассмотрены вероятностные модели поиска информации в информационном пространстве на примере университетской кампусной сети. Представлен принцип взаимоотношений между пользователями в рамках поисковых систем. Рассмотрены гипотезы о характере поведения пользователя в университетской кампусной сети.

Ключевые слова: университетские кампусные сети, информационный поиск, поддержка функционирования сети, поисковые системы.

Введение

Поддержка функционирования университетских кампусных сетей оправданно рассматриваются большинством ВУЗов Украины 4-го уровня аккредитации как один из важнейших вопросов, стоящий перед ними в настоящее время [6]. Университетская кампусная сеть (УКС) [7] – большая многосегментная компьютерная локальная сеть, развернутая на территории до нескольких квадратных километров, объединяющая локальные сети административных зданий, учебных корпусов, библиотек, общежитий, студенческих центров, спортивных сооружений, а также других зданий, входящих в состав ВУЗа.

Сеть подобного класса – это не только физическая инфраструктура, её планирование и построение на самом деле подразумевает разработку информационного пространства нового типа, позволяющего улучшить все аспекты функционирования ВУЗа [6].

1. Назначение и задачи университетских кампусных сетей

УКС предоставляет мощный динамический канал доступа к информации, а также способствует изменению механизма коммуникации кампуса. В своей ипостаси, сеть предоставляет широкий спектр возможностей, касающихся всех областей информационных технологий (вычисления, автоматизация, мультимедиа).

УКС [6] – один из стратегических аспектов в обеспечении высшего образования. Она занимает центральное место в образовании и совершенствует преподавание, обучение и научную деятельность путем предоставления преподавателям, студентам и прочему персоналу доступ к ресурсам кампуса. Кроме того, сеть снимает проблему информационной изоляции, тем самым предоставляя доступ к информации независимо от местонахождения преподавателя, студента и пр.

УКС улучшает атмосферу путем коммуникации между преподавателями, персоналом и студентами.

Многопрофильные учебные курсы могут быть спланированы и предоставлены от факультета к факультету и от факультета к студенту путем переговоров через электронную почту. Студенческие группы могут организовывать встречи, преподаватели могут распределить задания студентам, студенты могут принять участие в распределении заданий, библиотечные книги можно заказать по межбиблиотечному абонементу.

УКС поддерживает услуги по регистрации студентов на курсы, запрашиванию информации, планированию встреч и собраний, а также предоставлению электронных форм документов вместо бумажных вариантов. Помимо этого, сеть позволяет доносить информацию к абитуриентам, что полезно при наборе новых студентов.

УКС способствует эффективности управления путем предоставления оперативного доступа к данным заведения, улучшению принятых решений, повышению производительности труда и реинжинирингу рабочих процессов [6].

Поиск информации в УКС осуществляется через Web-интерфейс путем навигации по ссылкам, связанным с документами, и выявлением информации, соответствующей составленному условию поиска (далее запроса). Процесс поиска включает последовательность действий, суть которых сбор, обработка и визуализация информации, а именно это [8]:

- формирование запроса;
- идентификация информационных ресурсов, формально отвечающих условию поиска;
- визуализация информационных ресурсов;
- предоставление доступа к работе с ресурсами и оценка pertinентности поиска.

Определение 1. *Pertinentность* [8] найденной информации – соответствие выдаваемой поис-

ковой системой информации ожиданиям пользователя.

Стоит отметить, что релевантная модель оценки соответствия документа запросу практически не справляется с решением задач распознавания и поиска омонимов (грамматических и лексических), синонимов и многозначных слов из-за того, что в основу релевантной модели поиска заложен лингвистический подход и ряд оценочных синтетических критериев (таких как положение слов на странице). Очевидно, что вышеперечисленные языковые артефакты не могут быть распознаны без понимания смысла поискового запроса.

Определение 2. Релевантность [8] найденной информации - формальное соответствие пользовательского запроса к информации, выдаваемой поисковой системой.

В то же время пользователь должен получать не только информацию, максимально соответствующую критериям запроса, но и ту, которая отвечает предметной области и является актуальной.

2. Вероятностная модель информационного поиска

Имея только запрос, сеть не может точно определить релевантности документов. Для принятия решений в условиях неопределенности необходим математический аппарат теории вероятностей – вероятностный принцип ранжирования [9].

Суть принципа заключается в сортировке документов по оценённым вероятностям их релевантности запросу. Допустим, оценки релевантности – бинарные величины. Следовательно, документ может быть либо релевантным, либо не являться таковым. Таким образом, для каждого документа D и запроса q вводится случайная величина $R(D, q)$.

Ниже приведены основные положения принципа [10]:

- релевантность документа запросу не зависит от других документов в коллекции;
- пертинентность документа может зависеть от уже просмотренных документов пользователем;
- если поисковая система в ответ на каждый запрос ранжирует документы в порядке убывания их вероятности быть релевантными, и эта вероятность оценивается точно на основе доступных данных, то качество системы является наилучшим.

Для формального доказательства принципа применим следующий критерий качества – модель бинарных потерь: потери на документе D_i составляют $L(D_i) = 1$, если он был выдан, будучи не релевантным, либо не был выдан, будучи релевантным. В остальных случаях $L(D_i) = 0$. Тогда общие потери – сумма на всех документах коллекции.

Многие современные вероятностные модели поведения представляются в виде графических ве-

роятностных моделей. Эти модели представляют совместное распределение случайных величин, участвующих в вероятностной модели решаемой задачи, в виде графа. Отсутствие ребер кодируют условные независимости между величинами.

Преимущества таких графических моделей следующие [9]:

- визуальное представление структуры вероятностной модели является наглядным;
- важные свойства модели (условная независимость её случайных величин) сводятся к свойствам графа;
- сложные вычисления могут быть сформулированы в виде алгоритмов на графе.

3. Классификация графических моделей информационного поиска

Графическая вероятностная модель состоит из вершин и ребер, где вершины обозначают случайные величины $R(D, q)$, а ребра – связи между ними. Графическую модель принято различать по способу строения и виду самого графа (см. табл. 1).

Модели отличаются своими свойствами и распределением, тем не менее, основа алгоритма точного байесовского вывода (алгоритм передачи сообщений, в частности) удобно представима в виде фактор-графа [9]. Фактор-граф позволяет декомпозировать сложную функцию в произведение простых функций.

Таблица 1

Классификация графических вероятностных моделей

Класс	Основа
Байесовские сети доверия	направленные графы
Марковские случайные поля	ненаправленные графы
Алгебраические байесовские сети	графы

4. Социальное взаимодействие пользователей в поисковых системах

Разработке поисковых систем посвящено множество работ различной тематики, в частности, проектирование и разработка таких систем, которые учитывают социальные связи между пользователями [1–4].

Проблема поиска пользователями данных может быть решена путем создания единого, доступного всем хранилища информации, содержащего документы по тематике, необходимой пользователям [1]. Каждый пользователь, в свою очередь, может ограничивать поиск по определенным атрибутам при помощи фильтра, т.е. определять список свойств, характеризующий искомую информацию. Этот список может быть расширен и другими пользователями. Таким образом, точность идентификации любого документа, хранимого в системе, значительно повышается.

Следует отметить, что использовать подобную систему могут только пользователи, прошедшие

регистрацию. Также общее количество пользователей определяет эффективность такой системы.

Следующая проблема – мультиагентная организация обмена информацией между пользователями – и пути ее решения представлены в работе [2].

При помощи программных агентов система формирует персональный пользовательский профиль, учитывающая активность пользователя за время его работы с системой. Каждый пользователь организует собственную рабочую среду, путем создания профиля в системе (идентифицирующую его как участника данной системы). В таком профиле приведены персональные данные, рабочие проекты, а также ссылки на сопутствующие Web-документы.

Определение 3. Программный агент [8] - программа, которая вступает в отношение посредничества с пользователем или другой программой. Идея состоит в том, что агенты не запускаются непосредственно для решения задачи, а активизируются самостоятельно.

Система, анализируя данные профиля пользователя, формирует список ключевых слов, используемый в дальнейшем для поиска Web-документов по принципу совпадения ключевых слов со словами, встречаемых в текстовых документах. Результатом работы такой системы является определение рейтинга часто используемых источников информации для конкретного пользователя.

Обмен информацией между пользователями осуществляется исходя из близости предметных областей их профилей и выполняется благодаря отправке ссылок на Web-документы, отсутствующих у конкретного пользователя. Помимо отправки ссылок на источники ресурсов, предусматривается и отправка сообщений пользователю.

Поскольку решение проблемы мультиагентного обмена информации между пользователями [2] приводит к тому, что система становится подобием социальной, либо же кампусной сети, то следующая проблема [3] – обмен информацией между пользователями, которые объединены в кампусную сеть в рамках организации. Подобные системы функционируют по принципу опроса пользователей, необходимого для оценки эффективности установления ограничений на обмен информации между различными отделами организации. Определено, что для пользователей важным является получение информации в сжатые сроки, при этом он полностью доверяет источнику информации. Алгоритм поиска ресурса основывается, как правило, на базе имеющихся в общем доступе ресурсов и опыта разработчика системы.

Существует критерий поиска по предрасположенности некоторой группы людей к интересующей их информации [5] – социальный индекс (критерий оценки значимости документа). Другими словами,

документы могут быть проиндексированы не по ключевым словам, а по принадлежности к группе людей.

Социальный индекс документа для одного предпочтения участника сети определяется следующей формулой [5]:

$$SI = \sqrt{\frac{1}{2}MU^2 + \frac{1}{2}CU^2}, \quad (1)$$

где MU - средняя оценка привлекательности документа среди участников со схожими предпочтениями в диапазоне [0;10]; CU - соотношение общего числа документа к числу посещений участников с одним исследуемым интересом в диапазоне [0;1].

В основе формулы (1) лежит определение евклидова расстояния с двумя независимыми друг от друга параметрами MU и CU, а также весовыми коэффициентами (заданные относительно результатов экспертных оценок).

Принцип определения социального индекса заключается в выполнении следующих этапов [8]:

- получение/определение списка актуальных значений MU и CU для Web-документа, которые задаются другими участниками сети;
- выборка определенных значений MU и CU, учитывающие некоторые предпочтения определенной группы людей;
- расчет социального индекса документа;
- если оценка документа производится некоторой группой людей с различными предпочтениями, то социальный индекс рассчитывается для каждого участника сети по отдельности;
- тогда итоговое значение социального индекса определяется как сумма всех значений конкретных документов, которые соответствуют группе предпочтений текущего участника.

Применяя социальное индексирование как критерий поиска пертинентной информации, поисковая система будет формировать список ссылок на Web-документы в порядке соответствия предпочтениям пользователя.

5. Гипотезы о характере поведения пользователя в поисковых системах

Во время работы с УКС пользователь формирует запросы, предоставляющие ему доступ к документам D. В зависимости от роли (администратор, преподаватель, студент и т.п.) пользователь получает определенные права по взаимодействию с этими документами (общий случай – поиск и просмотр).

Взаимодействие с документами рассматривается в течение поисковой сессии. Сессия состоит из нескольких цепочек действий одного пользователя, разделённых небольшими интервалами во времени.

Оценки релевантности определяются для документов. Просмотр и выбор документа трактуются

как случайные события. Для конкретной сессии переменная R_i обозначает просмотр документа, показанный на позиции i ; S_i – выбор этой позиции.

Тогда вероятность просмотра документа на позиции i равна [9]:

$$P(R_i = 1),$$

и вероятность соответствующего выбора:

$$P(S_i = 1).$$

В табл. 2 представлены те гипотезы, на основании которых строятся все модели поведения.

Таблица 2

Классификация гипотез поведения пользователя

Название гипотезы	Описание
1) гипотеза просмотра	пользователь осуществляет выбор документа только после просмотра его описания. $P(S_i = 1) = P(R_i = 1)P(S_i = 1 R_i = 1)$
2) каскадная гипотеза	процесс просмотра описаний всегда начинается с первой позиции и строго линейен, т.е. просмотр описания документа D_i допустим только в случае, когда описание документа D_{i-1} было завершено. $P(R_1 = 1) = 1$ $P(R_{i+1} = 1 R_i = 0) = 0$
3) модель UBM (user browsing model)	вероятность просмотра описания документа D_i зависит от предыдущего выбранного документа D_{prev} и расстояния между позициями i и $prev$.

Результаты

Приведено определение университетской кампусной сети с определением её назначения и решаемых задач. Рассмотрены вероятностные модели поиска информации в информационном пространстве на примере университетской кампусной сети.

Представлен принцип взаимоотношений между пользователями в рамках поисковых систем.

Рассмотрены гипотезы о характере поведения пользователя в университетской кампусной сети.

Выводы

С целью повышения эффективности работы с университетской кампусной сетью следует искать технические решения, основываясь на вероятностном поиске информации и поведенческом анализе пользователя.

Список литературы

1. Губанов Д.А. Социальные сети: модели информационного влияния, управления и противоборства [Текст] / Д.А. Губанов, Д.А. Новиков, А.Г. Чартишвили. – М.: Физматлит, 2010 – 228 с.
2. Почанский О.М. Методы синтеза адаптивных web-страниц на основе интеллектуального анализа информационных ресурсов сети internet: Дис. ... канд. техн. наук: 05.13.23. – Х., 2012. – 168 с.

3. Обучающий материал “Технологии Semantic Web” [Электронный ресурс]. - Режим доступа к ресурсу: <http://habrahabr.ru/company/itits/blog/258405>.

4. Semantic Web [Электронный ресурс]. - Режим доступа к ресурсу: <http://semanticweb.org>.

5. Wikipedia: Semantic Web [Электронный ресурс]. - Режим доступа к ресурсу: https://en.wikipedia.org/wiki/Semantic_Web.

6. Thomas Moberg. Campus Network Strategies: A Small College Perspective [Электронный ресурс] / Thomas Moberg. – Режим доступа к ресурсу: <https://net.educause.edu/ir/library/html/cem/cem99/cem9916.html> – 10.12.2016 г. Загл. с экрана.

7. Margaret Rouse. Definition of Campus Network [Электронный ресурс] / Margaret Rouse. – Режим доступа к ресурсу: <http://searchsdn.techtarget.com/definition/campus-network> – 10.12.2016 г. Загл. с экрана.

8. Шостак И.В. Модель социального профиля как база предпочтений пользователя в процессе семантического поиска по Semantic Web / И.В. Шостак, А.А. Лысенко // Комунальне господарство міст: науково-технічний збірник Серія «Технічні науки та архітектура». – 2015. – Випуск 123. – С. 37-48.

9. Николенко С.И. Обзор моделей поведения пользователей для задачи ранжирования результатов поиска / С.И. Николенко, А.А. Фишков // Труды СПИИРАН. – 2012. – Вып. 22. – С. 139-175.

10. Robertson S.E. Probability ranking principle in IR / S.E. Robertson // Journal of Documentation. – 1977. – Vol. 33, N. 4. – P. 294-304.

Поступила в редколлегию 12.12.2016

Рецензент: д-р тех. наук, проф. Г.А. Кучук, Національний технічний університет «ХПІ», Харків.

АНАЛІЗ ПРОБЛЕМИ ПІДТРИМКИ ФУНКЦІОНУВАННЯ УНІВЕРСИТЕТСЬКИХ КАМПУСНИХ МЕРЕЖ

Є.А. Дружинін, І.В. Шостак, О.О. Лисенко

Приведено визначення університетської кампусової мережі з визначенням її призначення і вирішуваних завдань. Розглянуто ймовірнісні моделі пошуку інформації в інформаційному просторі на прикладі університетської кампусової мережі. Представлений принцип взаємовідносин між користувачами в рамках пошукових систем. Розглянуто гіпотези про характер поведінки користувача в університетській кампусової мережі.

Ключові слова: університетські кампусні мережі, інформаційний пошук, підтримка функціонування мережі, пошукові системи.

AN ANALYSIS OF THE PROBLEM OF CAMPUS NETWORKS FUNCTIONING SUPPORT

E.A. Druzhinin, I.V. Shostak, A.A. Lysenko

A definition of a campus network and solvable tasks by it are given. Probabilistic models of data searching in the campus networks are considered. A principle of user communication in the search engines is represented. Hypothesis of user behavior in the campus networks are considered.

Keywords: campus networks, data searching, network functioning supports, search engines.