

ИНТЕЛЛЕКТУАЛЬНАЯ ПОИСКОВАЯ СИСТЕМА НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ СОЦИАЛЬНОЙ СЕТИ

Е. А. Чирков

Бийский технологический институт (филиал)
Алтайского государственного технического университета им. И. И. Ползунова
г. Бийск

Среди популярных сервисов сети Интернет выделяются виртуальные социальные сети. В понятиях Интернета социальная сеть определяется как интерактивный многопользовательский веб-сайт, контент (информационное содержание) которого наполняется самими участниками сети. Основной способ организации пользователей внутри сети соответствует социологическому определению, согласно которому социальная сеть – это социальная структура (математически – граф), состоящая из группы узлов, которыми являются социальные объекты (люди или организации), и связей между ними (социальных взаимоотношений).

Существует множество вариаций социальных сетей по тематическим видам, одним из которых являются научные социальные сети. Научно-образовательная социальная сеть (НОСС) – тематическая социальная сеть, основными участниками которой являются ученые или люди (организации), связанные с наукой или имеющие к ней интерес.

Одной из главных черт способствующих объединению в группы и общению участников социальных сетей является система поиска контактов. Чаще всего людей в виртуальных сетях объединяет то же, что и в обычной жизни – общие интересы, работа. Для научных сетей важным является также и материал и опыт ученого. Для отображения деятельности и заслуг научного работника используются различные поля профиля, посвященные науке.

Цель исследования заключается в разработке интеллектуальной поисковой системы для научно-образовательной социальной сети. Качество сети рассматривается по легкости нахождения материала и контактов, что зависит от средств внутреннего поиска, т.е. от реализации поискового модуля.

Поиск – в широком смысле – стремление добиться чего-либо, найти что-либо или кого-либо, действия ищущего, направленные на получение нового: информации, данных, сведений, знаний и т.д.

Для поиска информации в Интернете широко используют поисковые системы и поисковые машины. Под поисковой системой понимают программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в Интернете. Поисковая машина – комплекс программ, предназначенный для поиска информации. Обычно является частью поисковой системы. Основными критериями качества работы поисковой машины являются релевантность (степень соответствия запроса и найденного, т.е. уместность результата), полнота базы, учёт морфологии языка. Разделяются на локальные и глобальные поисковые машины (системы):

- локальные предназначены для поиска информации по одному или нескольким сайтам, либо по локальной сети, являются внутренними поисковыми системами сайтов;

- глобальные осуществляют поиск информации по всей сети Интернет (поисковые машины систем поиска Google, Yahoo, Yandex и т.д.)

Поисковые машины осуществляют поиск информации различного типа, например текстов, видео, изображений, географических объектов, персональных данных и др.

При этом файлы, с которыми может работать поисковая машина, могут быть текстового (например .html, .htm, .txt, .doc, .rtf и т.д.), графического (.pdf, .gif, .png, .svg и т.д.) или мультимедийного форматов (видео, звук).

В последнее время появился новый тип поисковых движков, основанных на технологии RSS и XML-данных разного типа.

В поисковых машинах информация находится по поисковым запросам, которые задаются в виде части слова, набора слов или фразы, иногда используя расширенные возможности языка запросов поисковой системы.

Поисковые системы социальных сетей относятся к интеллектуальным благодаря «умному» поиску, основанному не на слепом поиске некоторой информации во всех доступных источниках, такие системы должны знать, где искать определенную информа-

цию, т.е. какие услуги использовать. Рассмотрим ситуацию, когда человек ищет книжный магазин: он использует "Желтые страницы" (упорядоченность информации по названию предоставляемых услуг) для области, в которой он проживает, находит список магазинов, выбирает из них один или несколько, и посещает их. Клиентский робот, далекий от интеллектуальных систем, шел бы во все магазины в области, спрашивая о книгах. В Сети, как и в реальной жизни, это неэффективно в малом и особенно в большом масштабе. Списки сайтов возвращаются отсортированными поисковыми системами по их посещаемости, определяется и значимость страницы. Анализ содержания страниц проводится согласно внутренним механизмам поиска поисковых машин.

Поисковые машины выполняют несколько функций:

- автоматический и ручной поиск ссылок на страницы и другие документы сайтов;
- классификация документов сайта с проведением индексации (извлечение, преобразование и сохранение важной информации для поиска в БД поисковой машины), поиском по БД проиндексированных документов, построением и обучением классификатора или кластеризацией документов.

Поиск документов осуществляется по соответствию поисковому запросу с ранжированием документов в соответствии с их релевантностью поисковым запросам.

Классификация документов заключается в отнесении документа к одной из нескольких категорий на основании содержания документа. Классификация создается вручную, либо автоматически с помощью созданного вручную набора правил, либо автоматически с применением методов машинного обучения.

Целью кластеризации документов является автоматическое выявление групп семантически похожих документов среди заданного фиксированного множества документов. Следует отличать классификацию текстов от кластеризации, в последнем случае тексты также группируются по некоторым критериям, но заранее заданные категории отсутствуют.

Поисковые системы типа Google, Yahoo, Rambler и Yandex превосходят внутренние поисковые машины сайтов, даже таких крупных, как социальные сети Facebook, Twitter, MySpace, «В Контакте», проверкой орфографии, преобразованием (неправильно) набранного транслита, специально подбираемой контекстной рекламой, ответами на об-

щие вопросы (или даже решением формул и построением графиков, как в pigma.ru). Но социальные сети предоставляют также инструменты поиска людей и групп, входящих в их состав.

Зачастую пользователю при входе в новую и особенно крупную социальную сеть сложно найти в её многообразии участников свой круг общения, именно в этом случае важно насколько качественно продуман и реализован поиск участников сети. Классическими стали поиск по схожим интересам и поиск одноклассников. Поиск по интересам выводит обычно большой список пользователей (или групп), но обычно присутствует ранжирование по посещаемости популярности и т.п. Поиск одноклассников сводится к выбору пользователей по городу, школе (также страны, класса, вуза и группы), и годам обучения (или возраста и пола участников). Для поиска коллег можно использовать специальные инструменты и организации и т.д. В социальных сетях есть также помощь в поисках конкретных людей: поиск по e-mail, ICQ, Skype, Windows Live Messenger и т.д. Другие виды поиска людей в социальных сетях к интеллектуальным не относятся.

Среди социальных сетей можно отметить поиск контактов:

- «Одноклассники» (и Classmates) – с поиском школьных знакомых;
- Facebook.com – поиск уже знакомых друзей в сети с использованием номера ICQ, e-mail и т.д.;
- Academia.edu – представлен визуальный поиск в виде графической древовидной структуры преподавателей академий и вузов.

При создании социального сайта важен выбор движка сайта – CMS (система управления веб-контентом). Если разработчики сети не создают движок сети с нуля, то пользуются готовыми бесплатными или платными CMS. CMS поддерживающие функционал социальной сети обычно обладают внутренними подсистемами поиска по содержанию сайта и поиска пользователей.

На примере поисковых модулей CMS Drupal и InstantCMS можно найти способ создания оптимальной поисковой системы.

Основное направление поиска в Drupal связано с нахождением материала. В Drupal поддерживается режим индексирования с настраиваемыми параметрами минимальной длины слова для индексации и количеству индексируемых элементов. Сортировка поддерживается ранжированием результатов

поиска по нескольким параметрам: релевантность ключевого слова, дате отправления, числу комментариев и просмотров.

Для CMS Drupal свойственна таксономия содержания, т.е. поиск осуществляется по категориям (похожий на поиск в веб-форумах). Такой поиск является достаточным для нахождения материала, но неудобен для поиска контактов.

В InstantCMS помимо обычного индексированного поиска по содержанию сайта присутствует также поиск пользователей и поиск по тегам и отображение ссылок на похожий материал (свойственный для фолксономии). Для поиска пользователя присутствует специальный блок, в который вносятся параметры поиска: диапазон возраста, пол, интересы, город. При запросе происходит обращение к БД пользователей по одному или нескольким столбцам (в профиле эти столбцы отображаются как анкетные поля для заполнения), далее происходит считывание данных по всем пользователям и вывод результатов по релевантности запроса. Данный модуль можно расширить добавлением дополнительных полей для запроса и расширением условий (например, высокий рейтинг в сети или положительная «карма»).

При разработке интеллектуальной поисковой системы в ИС научно-образовательная социальная сеть необходимо учесть, что поиск пользователем может осуществляться по 3 направлениям:

- поиск материала;
- поиск пользователей;
- поиск в категориях и тегах.

Для удобства поиска по категориям в случае большого числа результатов можно использовать различные фильтры, как алфа-

витный указатель, популярность, рейтинг, дату публикации (посещения, регистрации) и т.д. Теговые облака и теговый поиск помогают найти популярный или похожий материал без прямого использования модуля поиска.

Подсистемы поиска по данным профиля, теговую фолксономию, списки и алфавитные указатели. Для этого можно выбрать средства CMS Drupal, стабильную в работе с мощным административным инструментом и обширным списком дополнительных модулей, решающих большинство задач при построении функционала ИС НОСС.

СПИСОК ЛИТЕРАТУРЫ

1. Социальная сеть / Материал из Википедии ru.wikipedia.org
2. Поисковая система / Материал из Википедии ru.wikipedia.org
3. Поисковая машина/ Материал из Википедии ru.wikipedia.org
4. Поисковые системы в сети Интернет citforum.ru/internet/search/searchsystems.shtml
5. Жёлтые страницы / Материал из Википедии ru.wikipedia.org
6. Nigma научилась чертить графики habrahabr.ru/blogs/nigma/94569/
7. Поисковик Nigma.ru. Математика с учетом области допустимых значений habrahabr.ru/blogs/nigma/78531/
8. Об информационном поиске, нахождении оптимальных путей просмотра результатов поиска habrahabr.ru/blogs/search_engines/76301
9. Фолксономия и/или таксономия blog.vityasev.ru/2007/02/06/folksonomy-and-or-taxonomy/
10. Таксономия и фолксономия iskariot.ru/development/organization-of-content/
11. CMS Drupal. Search: an internal site search system drupal.org/handbook/modules/search/
12. Сайт InstantCMS www.instantcms.ru