

«Сейфуллин оқулары – 18: « Жастар және ғылым – болашаққа көзқарас» халықаралық ғылыми -практикалық конференция материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 18: « Молодежь и наука – взгляд в будущее» - 2022.- Т.І, Ч.ІV. - С. 19-21

АНАЛИЗ ИНСТРУМЕНТОВ С ОТКРЫТЫМ ИСХОДНЫМ КОДОМ ДЛЯ СБОРА И ВИЗУАЛИЗАЦИИ КОНТЕНТА СОЦИАЛЬНЫХ СЕТЕЙ

Жұматай Б.Ә.

Казахский агротехнический университет имени С.Сейфуллина, город Нур-Султан

Введение

Данные и информация из социальных сетей становятся реальной заботой для многих сторон, таких как страны и организации, для целей анализа, стратегии и принятия решений. Выбор лучшего инструмента с открытым исходным кодом для сбора и визуализации данных из социальных сетей представляет собой большую проблему. В этой статье мы предлагаем интегрированный подход к выбору наиболее подходящих инструментов для сбора и визуализации информации, доступной в социальных сетях. Предлагаемый подход состоит из двух этапов, а именно оценки и выбора. Текущая практика оценки включает в себя мнения пользователей-экспертов об инструменте, чтобы оценить его важность на основе удовлетворенности пользователей, которая измеряется эффективностью инструмента для поиска необходимых результатов. Фаза оценки предлагаемой методики сочетает в себе текущую практику с весом важности, где последний был рассчитан с помощью статистических методов. Фаза выбора разработана на основе известной техники PROMETHEE (хорошо продуманная система поддержки принятия решения).

Онлайновые социальные сети (ОСС) становятся одним из крупнейших источников огромного количества данных. Этот источник содержит мысли, обсуждения и дебаты, выраженные в публичных социальных беседах, которые могут быть использованы не только в качестве важного компонента процесса принятия решений, но и в промышленных, академических и политических решениях. Facebook, YouTube и Twitter (ВКонтакте- в странах СНГ) являются самыми популярными онлайн-социальными сетями. Они широко распространены среди широких слоев населения. Facebook объявил о 1,39 млрд ежемесячных активных пользователей в декабре 2020 года. Twitter насчитывает более 500 миллионов пользователей, а YouTube - более одного миллиарда пользователей. В этих сетях постоянно создается и распространяется большое количество пользовательского контента. Существует острая необходимость в получении этой информации из

социальных сетей, ее хранении, анализе и визуализации, чтобы можно было извлечь из нее пользу в нескольких аспектах.

Сбор, хранение, обработка, анализ и визуализация данных социальных сетей являются сложными задачами. Из-за огромного количества данных в социальных сетях очень трудно решить эти задачи без сопутствующего инструмента.

Существующие исследования используют некоторые из этих инструментов и библиотек для проведения своих экспериментов. В большинстве из них упоминается использование инструмента для поиска твитов, профилей пользователей или атрибутов видео на YouTube.

В этой статье мы предложили интегрированный подход, который сначала поможет оценить существующие инструменты и библиотеки-кандидаты, а затем выбрать лучший из них, который отвечает требованиям и потребностям пользователей для сбора, хранения, анализа и визуализации содержимого OSNs (Open Systems Network Support). Мы рассматривали только инструменты и библиотеки для сбора и визуализации [1, с.74]. Для оценки инструментов мы провели небольшой опрос. В опросе приняли участие двенадцать человек, которые работают IT-специалистами в нашем городе Нур-Султан и специализирующихся как на социальных сетях, так и на CASE-инструментах. Они помогли определить относительную важность (вес) каждого из предложенных критериев. Наш основной вклад в данную работу заключается в следующем:

- Мы разработали подход к подсчету баллов на основе определенного веса важности и текущей практики использования каждого инструмента.

- Мы рассмотрели более 25 инструментов с открытым исходным кодом и отобрали 12, в соответствии с их популярностью, количеством загрузок и историей обновлений.

- Насколько нам известно, в нашей работе впервые предложен интегрированный подход к выбору подходящей структуры и инструмента для сбора и визуализации данных.

Сбор данных - это концепция сбора данных в хранилище с целью анализа, который приводит к получению полезной информации или знаний. Существует множество инструментов, которые можно использовать для сбора данных из приложений социальных сетей, называемых "краулеры". Краулер - это программа, которая перемещается по информационному пространству WWW, следуя гипертекстовым ссылкам и извлекая веб-документы по стандартному протоколу HTTP.

Визуализация данных - это техника представления проанализированных данных в коммуникативном, понятном и эффективном виде. Визуализация данных имеет набор форматов или способов отображения необходимых результатов [2, с.96]. Эти способы включают приборные панели, диаграммы, радары и датчики, которые дают пользователю возможность выбрать любой формат или схему, подходящую и предпочтительную для него.

Методы MCDM (многокритериальный анализ) помогают лицам, принимающим решения, анализировать различные критерии выбора инструментов, оценивать альтернативные варианты CASE-средств и делать желаемый выбор инструментов. Для выбора было предложено множество методов MCDM, таких как процесс аналитической иерархии (АИР) и метод организации ранжирования предпочтений для оценки обогащения (PROMETHEE).

После сбора результатов мы рассчитали разницу между важностью и текущей практикой выбранного элемента. Уравнение (1) рассчитывает средний вес сбора $W[i] = L_{j=1}^a w [i, j]/a$ данных i для оценок: (1)

Где a означает количество инструментов и знаний, которые рассматривались в вопросе, w означает вес важности, c означает текущую практику. Уравнение (2) рассчитывает среднее значение текущей практики сбора данных i :

$$C[i] = \frac{1}{a} \sum_{j=1}^a c[i, j] \quad (2)$$

Уравнение (3) рассчитывает коллективный взвешенный показатель реализации V для N элементов:

$$V = \left\{ \frac{\sum_{i=1}^N W[i] * c[i]}{\sum_{i=1}^N W[i] * (5)} \right\} * 100 \quad (3)$$

Последний шаг - извлечение оценок важности и текущей практики, чтобы получить самый высокий и самый низкий балл по предложенным элементам; мы применили уравнение (3). В результате, если процент высокий, это означает, что данный элемент учитывается в текущем инструменте, с другой стороны, если процент низкий, это означает, что элемент является важным и не учитывается в текущей практике.

Rank	action		Phi	Phi+	Phi-
1	Gephi	<input checked="" type="checkbox"/>	0,2167	0,4167	0,2000
2	JUNG	<input checked="" type="checkbox"/>	0,0500	0,2667	0,2167
3	D3	<input checked="" type="checkbox"/>	0,0000	0,2000	0,2000
4	amCharts	<input checked="" type="checkbox"/>	0,0000	0,2333	0,2333
5	Kap-lab	<input checked="" type="checkbox"/>	-0,0333	0,2000	0,2333
6	Prefuse	<input checked="" type="checkbox"/>	-0,2333	0,0833	0,3167

За последние несколько лет заметно возросло использование онлайн-социальных сетей отдельными лицами, странами и организациями для анализа, разработки стратегий и принятия решений.

Выбор лучшего инструмента для социальных медиа с открытым исходным кодом является одной из самых больших проблем для сбора и визуализации данных [3, с.114]. В данной работе мы предлагаем интегрированный подход к выбору наиболее подходящих инструментов для сбора и визуализации информации в социальных сетях. Этап оценки предлагаемой методики представляет собой сочетание текущей практики со статистическими методами, а этап выбора основан на известной методике PROMETHEE с усовершенствованием. Результаты показывают, что предложенный подход является перспективной методикой выбора инструмента для анализа данных социальных сетей. В качестве будущей работы мы можем расширить эту работу для хранения и добычи инструментов анализа в онлайн-социальной сети.

Список использованной литературы

1. Винник В. Д. Социальные сети как феномен организации общества: сущность и подходы к использованию и мониторингу // Философия науки. - 2012. - №4 (55). - С. 110-126.
2. Патаракин Е. Д. Педагогический дизайн социальной сети Scratch // Образовательные технологии и общество (Educational Technology & Society). - 2013. - Т.16. - № 2. - С. 505-528.
3. Скопина, И. Маркетинговые коммуникации в социальных web-ресурсах / И. Скопина, Ю. Бакланова, А. Пименова // Маркетинг. – 2008. - № 5. – С. 43 – 50.