

## АВТОМАТИЗИРОВАННАЯ СИСТЕМА КРАУДСОРСИНГА ДАННЫХ

*Кудин Н.И., Сабадаш М.А.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Калугина М.А. – канд. физ.-мат. наук, доцент*

В настоящее время системы краудсорсинга активно развиваются, поскольку данный способ сбора информации является эффективным как для рынка Западной Европы, так и для стран СНГ. Данный подход открывает много возможностей для создания уникального сервиса обработки данных, основанных на запросах пользователя.

Современные технологии развиваются ежедневно с высокой скоростью. Одной из проблем, которую приходится решать компаниям в различных сферах, является поиск инновационных идей для дальнейшего развития [1].

Как правило, крупные компании используют высококвалифицированных сотрудников для решения данной проблемы. Но, к сожалению, небольшие компании, как правило, имеют ограниченный бюджет, тем самым не могут себе позволить реализацию инновационных задач в достаточном объеме.

Примером больших компаний, которые находятся в постоянном поиске улучшений условий бизнеса, могут служить British Gas, British Airways, Virgin Trains и другие. Такие крупные «игроки» всегда заботятся о перспективности предлагаемого решения и скорости воплощения идеи в жизнь, так как реализация какого-либо проекта должна выполняться в такие сроки, чтобы идея сохранила свою актуальность [2].

В качестве решения данной проблемы предлагается использование высоконагруженных систем краудсорсинга, которые позволяют в краткие сроки производить анализ профессиональной активности работников компании и выявлять наиболее перспективные идеи для поставленной задачи. Кроме того, данные системы могут организовать отклик сотрудников на проблемные места компании, уменьшить расходы на высококвалифицированные кадры, исключить необходимость постоянного аудита ключевых отделов компании. Такие системы в режиме реального времени дают возможность производить синхронизацию с открытыми источниками данных (например, корпоративные социальные сети). Собрав пакет данных, можно выполнить анализ эффективности работы компании и определить ее слабые звенья [3].

На рисунке 1 представлена схема стандартного процесса краудсорсинга идей.

Примером системы краудсорсинга является Yahoo!Answers [4]. Сервис вопросов и ответов Yahoo!Answers запущен еще в 2005 году. Благодаря сервису любой пользователь может создать вопрос, а все добровольцы из сообщества, у которых есть несколько баллов за предыдущую активность, — попытаться ответить на него. Если хотя бы один из ответов помог решить проблему, пользователь, задававший вопрос, может отметить его как лучший и присвоить ему оценку, которая повышает авторитет пользователя. Большая англоязычная аудитория данного сервиса делает его перспективным для монетизации сервиса.

Предлагаемая нами автоматизированная система краудсорсинга данных предполагает собой абстракцию между источниками данных и непосредственно системой анализа данных. Таким образом, пользователь системы может самостоятельно выбрать необходимые ему источники данных, которые должны быть проанализированы. В свою очередь, система выполняет анализ и отображение данных в виде графиков. Данный способ визуализации помогает пользователям ускорить процесс определения тенденций в источниках данных. Также в системе предусмотрена агрегация данных из разных областей, что поможет определять общий вектор направленности данных.

Особенностью решения является глубокая интеграция сервисов с корпоративными социальными сетями, с помощью которого пользователи смогут решать поставленные задачи. Используя данный подход, целевая аудитория будет расширена многократно.

Алгоритм работы синхронизации является одной из ключевых частей программного продукта. В самом начале работы алгоритма система выбирает все доступные и активные компании в системе. Затем синхронизатор производит подключение к источнику данных (например Yammer). Следующим этапом является постраничная обработка данных в соответствии с настройками компании. Тут же будет работать проверка на дубликаты идей. После того, как идея была получена, будет произведена проверка по таким параметрам как:

- тип рабочей сети;
- группа идеи;
- тема идеи;
- дата создания.

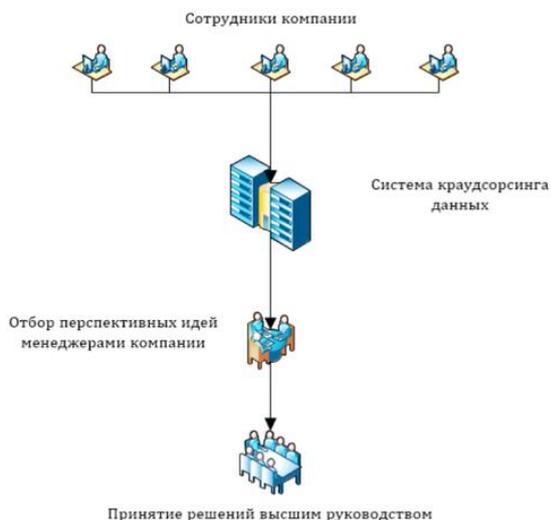


Рисунок 1 – Схема процесса краудсорсинга идей

Если идея будет удовлетворять всем параметрам – идея будет сохранена в базе данных, иначе будет произведен переход к следующей идее в очереди. Когда очередь станет пустой, механизм синхронизации закончит свою работу. Разработанный алгоритм представляет собой агрегацию более низкоуровневых алгоритмов, которые применяются для сбора данных в зависимости от источника. Например, каждая корпоративная социальная сеть или же мессенджер, имеют свой API, в котором модели данных различаются. Сложность задачи состоит в том, чтобы произвести агрегацию данных и представить данные в едином виде.

Передовыми технологиями в данной области являются сверхпроизводительные сервисы Azure [5], служащие для распределения нагрузки в системе. Так, для анализа данных имеет смысл использовать Text Analysis - сервис Azure для качественной оценки данных. Также данный сервис позволяет выделить ключевые слова для каждой выделенной сущности. Полученные данные могут быть использованы при осуществлении аналитики данных.

Кроме того, для области синхронизации с источниками данных актуально применение микро-сервисной архитектуры с элементами событийно-ориентированной модели для максимальной оптимизации приложения. Данная оптимизация касается как скорости работы системы, так и объемов потребляемых ресурсов. К примеру, использование Azure Service Bus позволяет воспользоваться отложенным выполнением задач в системе, а также построить максимально гибкую систему обработки информации. Кроме того, шина сообщений позволяет создать систему с поддержкой распределенной нагрузки, которая будет возникать в случае большого количества данных в источниках.

Применение Azure Service Fabric позволяет добиться высокой производительности системы и гибкого увеличения ресурсов по мере необходимости. Предполагаемая нагрузка системы, как правило, высока ввиду большого количества данных в источнике.

Таким образом, решение проблемы оценки степени инновационности идей можно выполнить с помощью современных технологий. Предлагаемая автоматизированная программная система краудсорсинга идей облегчит поиск перспективных направлений развития компании, аудит рабочих процессов в отделах, а также будет способствовать вовлеченности сотрудников в процесс продвижения компании на рынке конкурентов.

**Список использованных источников:**

1. Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business - Jeff Howe, 2014.
2. Тапскотт Д, Энтони Д. Уильямс. Викиномика. Как массовое сотрудничество изменяет все / Тапскотт Д, Энтони Д. Уильямс. – BestBusinessBooks. – 2009. – 223 с.
3. Crowdsourcing for Dummies – David Alan Grig, 2013.
4. Краудсорсинг: основные площадки [Электронный ресурс]. - <http://www.towave.ru/pub/kraudsorsing-osnovnye-ploshchadki-v-rossii-i-ikh-analogi-na-zapade.html> Дата доступа: 05.04.2017.
5. Azure [Электронный ресурс]. - <https://azure.microsoft.com/en-us/services/> Дата доступа: 05.04.2017.