

УДК 004.89

*Брещук Д. А.*

*Научный руководитель: к. т. н., доцент кафедры ИИТ Крапивин Ю. Б.*

## МЕТОДЫ АВТОМАТИЧЕСКОЙ АГРЕГАЦИИ НОВОСТНЫХ СООБЩЕНИЙ

Целью данной работы является анализ наиболее популярных источников новостной информации подходящих для решения задач их автоматизированной обработки.

В сегодняшнем высокотехнологичном мире все больше увеличивается значимость информации и связанных с ней информационных технологий. Динамика роста объема информации такова, что ее количество удваивается каждые 20 месяцев.

В настоящее время существует множество новостных источников информации, требующих их оперативной и качественной обработки в целях решения практических задач. К таким задачам можно отнести получение качественной и проверенной информации по определенной тематике, систематизация сведений из заданной предметной области, классификация сообщений в соответствии требованиями пользователя, автоматический sentiment-анализ, создание роботизированной ленты новостей и т. п.

Важной составляющей систем, решающих указанные задачи, является достоверный и актуальный источник информации. На данный момент нет единых реестров или баз данных, в связи с чем информацию приходится собирать из различных источников, которые, как правило, не предназначены для программного использования (например, веб-сайты, сторонние системы). Широко распространенной практикой является применение формата JSON (JavaScript Object Notation) – общего формата представления значений и объектов для хранения и обмена новостной информацией. Некоторые источники предоставляют API (application programming interface), благодаря которому сторонние разработчики могут получить к ним доступ. Здесь стоит учитывать, что такое API не всегда является открытым и может предполагать регистрацию и плату за использование.

### **News API**

News API [1] предоставляет сервис для поиска новостей в 54 странах, а это больше чем 30 000 новостных агентств. Данный обширный охват источников предоставляет пользователю множество возможностей для точного анализа новостей. Например, вы можете узнать какие новости сейчас в тренде или определить, что пишут об определенной компании в разных странах.

Сервис предоставляет API. Поддерживается возможность поиска опубликованных статей по ключевым словам или фразам, языкам, названиям источников публикации, датам публикации и доменным именам источников публикации, отсортировать результаты по дате публикации, популярности источника публикации или релевантности ключевого слова поиска. API распространяется по условно-бесплатной форме (бесплатно его можно использовать только для тестирования перед покупкой) [2]. Также у бесплатной версии есть и ограниче-

ния, основными из которых являются: ограничение по количеству запросов в день, задержка доступа к новостному сообщению на 1 час, а также ограничения на поиск новостей только месячной давности (в платной версии – трехлетней давности).

В качестве ответа может быть получен JSON-файл, содержащий в себе всю необходимую информацию, такую как автор, дата публикации, заголовок и т. д. (пример ответа на рисунке 1). [3] В качестве недостатка можно отметить, что сервис не предоставляет полный текст новости.

```
{
  "status": "ok",
  "totalResults": 9186,
  "articles": [
    { //Первый ответ
      "source": {
        "id": null,
        "name": "Observer"
      },
      "author": "Sissi Cao, Sissi Cao",
      "title": "Elon Musk Says a Pong-Playing Monkey Proves Neuralink Works, But He's Promised A Lot",
      "description": "However, Elon Musk has a history of missing product timelines he sets for himself.",
      "url": "https://observer.com/2021/04/elon-musk-neuralink-monkey-video-game-tech-neurotech-human-application/",
      "urlToImage": "https://observer.com/wp-content/uploads/sites/2/2021/04/neuralink.jpg?quality=80&strip",
      "publishedAt": "2021-04-09T16:08:22Z",
      "content": "Elon Musk's controversial neurotech startup Neuralink just made a major breakthrough in its brain-machine interface."
    },
    { //Второй ответ
      .....
    }
  ]
}
```

*Рисунок 1 – Пример ответа от News API: с запросом на все статьи о Тесла в течении месяца, отсортированные по времени*

## **Bing News API**

Bing News API [4] – это система, основанная на технологии искусственного интеллекта, которая является частью поисковой системы Microsoft Bing, обрабатывающей миллиарды глобальных поисковых запросов, предоставляя доступ к новостным статьям со всего мира.

Основные возможности API: отправка поисковых запросов на платформу и получение исчерпывающих результатов последних новостей, интересных статей, актуальных тем и изображений со всего мира. Результаты запроса могут быть отфильтрованы по теме, сведениям об издателе, местным новостям и другим доступным для поиска метаданным. При поиске новостей существует возможность поиска по одной из трех категорий:

- **Новости в категории** – возвращает самые популярные новостные статьи по категориям. Например, возможно запросить самые популярные статьи о спорте или развлечениях.

- **Поиск** – возвращает новостные статьи на основе поискового запроса пользователя. Если поисковый запрос пуст, вызов возвращает самые популярные новостные статьи.

- **Популярные темы** – возвращает популярные темы новостей, которые в настоящее время популярны в социальных сетях. Доступно только для США [2].

При этом Microsoft предлагает бесплатную версию и несколько платных тарифных планов. В платных версиях меняется количество запросов в секунду и количество запросов в месяц (чем дороже, тем больше запросов можно совершить).

Как и предыдущий сервис, Bing News API возвращает всю необходимую информацию (включая ссылку на видео и связанную с ним информацию) в виде JSON-файла, за исключением полного текста новости, пример ответа представлен на рисунке 2 [3].

```
{
  "_type": "NewsArticle",
  "name": "Tesla's AI is about to get better, and not just for Full Self-Driving",
  "url": "https://electrek.co/2021/04/12/tesla-ait-better-not-just-for-full-self-driving/",
  "image": {...},
  "description": "Tesla's use of artificial intelligence in its electric vehicles i..",
  "provider": [...],
  "datePublished": "2021-04-12T10:25:00.000000Z",
  "video": {
    //Информация о видео файле
    "_type": "VideoObject",
    "name": "Tesla's AI is about to get better and not just for Full Self-Driving - Electrek",
    "thumbnailUrl": "https://www.bing.com/th?id=OVF.hrFKR6ZpIPc8TnI1%2F5TK3g&pid=News",
    "thumbnail": {...},
  },
  "category": "ScienceAndTechnology",
}
```

*Рисунок 2 – Пример ответа от Bing News API: - с запросом на поиск новостей о Тесла*

### **Google News API**

Google News API [5] – это популярная система, которая предоставляет самую свежую новостную информацию, собранную из многочисленных источников по всему миру на 35 языках, не смотря на то, что поддержка прекращена.

API позволяет интегрировать результаты поиска новостей в приложение или веб-страницы, предоставляя следующие варианты запросов:

- Заголовки тем – поиск по определенной теме.
- Главные заголовки – последние популярные статьи.
- Поиск – поиск по определенному слову или фразе.
- Поиск по источнику – поиск по определенному слову или фразе в данном источнике.
- Географические заголовки – поиск по географическому положению.

Кроме того, присутствует возможность использовать его для отображения тем, заголовков, популярных историй, URL-адресов и других новостей из результатов поиска Google.

API можно использовать бесплатно с одним ограничением в 3 запроса в час. Поддерживается несколько платных версий, которые отличаются количеством запросов в секунду и количеством запросов в месяц.

Ответом от API является JSON-файл, содержащий заголовок, ссылку на новость, дату публикации, id-источник, похожие статьи. Пример ответа приведен на рисунке 3 [3].

```

"articles": [
  0: {
    "id": "52781509581326"
    "title": "Buy Tesla Stock Because It Could Be the Next Apple. Here's How. - Barron's"
    "link": "https://www.barrons.com/articles/buy-tesla-because-it-could-be-the-next-apple-heres-how-51618232801"
    "published": "Mon, 12 Apr 2021 13:06:00 GMT"
    "sub_articles": [
      0: {
        "url": "https://www.barrons.com/articles/buy-tesla-because-it-could-be-the-next-apple-heres-how-51618232801"
        "title": "Buy Tesla Stock Because It Could Be the Next Apple. Here's How."
        "publisher": "Barron's"
      }
    ]
    "source": {...}
  },

```

**Рисунок 3 – Пример ответа от Google News API: с запросом на поиск новостей о Тесла**

Таким образом, любая система может быть ориентирована на работу с одним из представленных выше сервисов в зависимости от тематической направленности анализируемой информации.

#### **Список цитированных источников**

1. News API [Электронный ресурс]. – Режим доступа: <https://newsapi.org>. – Дата доступа: 08.04.2021.
2. Список новостных API [Электронный ресурс]. – Режим доступа: <https://medium.com/rakuten-rapidapi/top-10-best-news-apis-google-news-bloomberg-bing-news-and-more-bbf3e46af6>. – Дата доступа: 12.04.2021.
3. Доступ к API [Электронный ресурс]. – Режим доступа: <https://rapidapi.com/blog/rapidapi-featured-news-apis>. – Дата доступа: 05.04.2021.
4. Bing News API [Электронный ресурс]. – Режим доступа: <https://www.microsoft.com/en-us/bing/apis/bing-news-search-api>. – Дата доступа: 11.04.2021.
5. Google News API [Электронный ресурс]. – Режим доступа: <https://rapidapi.com/newscatcher-api-newscatcher-api-default/api/google-news>. – Дата доступа: 05.04.2021.

УДК 004.89

*Ильяшевич Д. А., Супрунчук П. Д.*

*Научный руководитель: ст. преподаватель Клопоцкий А. А.*

## **МОБИЛЬНЫЙ ПОМОЩНИК ДЛЯ СТУДЕНТОВ БРЕСТСКОГО ГОСУДАРСТВЕННОГО ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА**

При широком распространении электронных средств коммуникации является актуальной задача разработки различных приложений. В данном проекте разработано мобильное приложение «БрГТУ Ассистент», позволяющее пользователям (т. е. студентам) быстро получать достоверную информацию об орга-