

Организация мониторинга
открытых источников в
интернете

Мониторинг открытых источников

систематический сбор, обработка и классификация информации из общедоступных источников (СМИ, научных публикаций, баз данных и т. д.)
чаще всего – это отслеживание определенной темы в информационном потоке



Изучение тенденций
в предметной
области
*для накопления
информации и
последующей подготовки
справок, аналитических
документов*



Отслеживание
медийной
активности
объектов
*мониторинг
публичных
действий, заявлений
интересующих лиц*



Изучение
информационной
повестки (СМИ,
региона)
*мониторинги могут
использоваться для
анализа
информационной,
политической
обстановки*



Изучение имиджа
(персоны, организации)
*как воспринимается
та или иная фигура
или организация в
публичном
пространстве*



Определение
информационного поля
конкретной темы
*как СМИ, влиятельные
фигуры, массовая
аудитория реагируют на
событие или проблему*

Традиционные объекты интереса при мониторинге

1



Личность

Лидеры мнений, эксперты (политики, бизнесмены, журналисты, ученые и т. д.)

2



Организация
(совокупность)

«Факультет журналистики БГУ», «Ведущие СМИ Латинской Америки»

3



Социальная
группа

«Журналисты негосударственных СМИ», «Рабочие агропромышленного комплекса»

4



Тематическая
область

Сфера деятельности (образование, здравоохранение), тематический сегмент (социальные медиа, цифровая журналистика)

5



Географический
регион

«Браславский район», «Минск», «Аргентина»

Интернет как источник информации

Видимая сеть (Surface Web)

4%

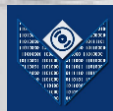
легкодоступна,
индексируется
поисковыми системами
(5,55 млрд страниц на
сентябрь 2020 года)



Глубокая сеть (Deep Web)

96%

- страницы, которые трудно индексировать (запрет, динамичный контент, нет ссылок)
- доступ к сайту по паролю или после регистрации
- локальные сети

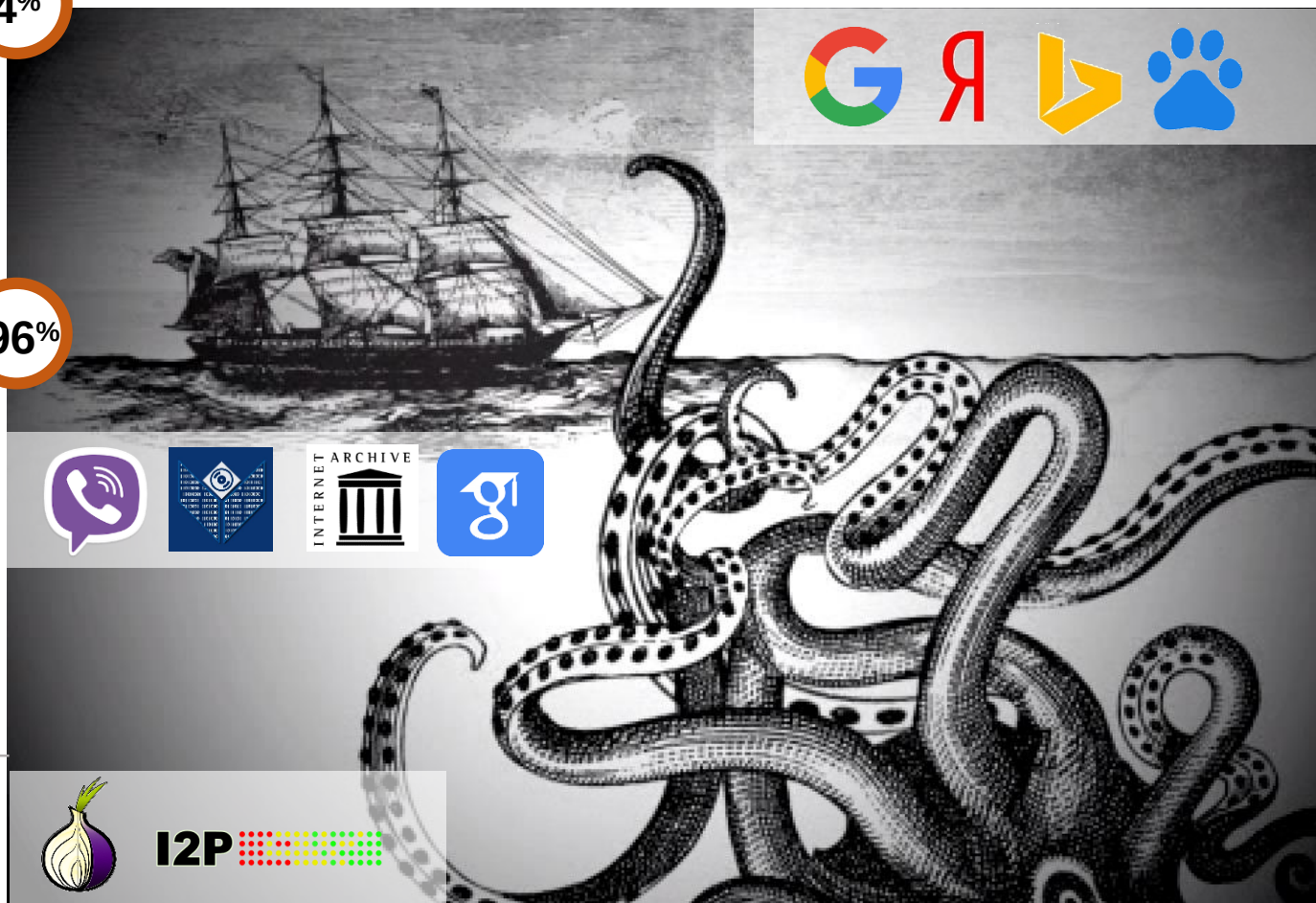


Теневая сеть (Dark Web)

анонимная
файлообменная сеть;
нужны спец. программы
для подключения



I2P



Инструменты для сбора информации из открытых источников в интернете

Добыча информации

- Поисковые машины
(Google, DuckDuckgo, SurfWax)
- Социальные медиа
(Facebook, Вконтакте)
- Каталоги, библиотеки
(Unicat.nlg.by)
- Архивы
(Wayback Machine)
- Файлообменные системы
(RuTracker, VK)
- Экспертные сервисы
(Quora)

+ объявления / вакансии, тематические сайты,
базы данных онлайн, массивы СМИ, энциклопедии...

Мониторинг

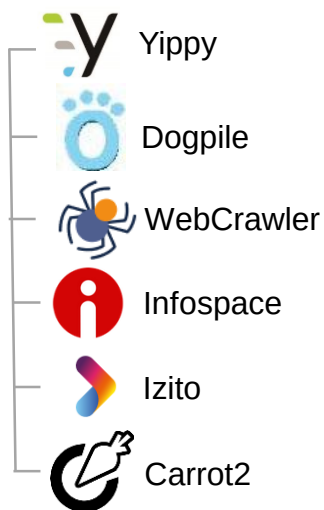
- Сервисы упоминаний
(Google Alerts)
- RSS-агрегаторы
(Feedly, QuiteRSS)
- E-mail-рассылки
- Системы мониторинга сайтов
(WebSiteWatcher)
- Сервисы мониторинга
социальных медиа
(IQ Buzz, Babkee)
- Информационно-
аналитические системы
(SiteSputnik)



Поисковые машины

Метапоисковые

ищут сразу по
нескольким системам



■ как правило, охватывают
Google, Bing, Yahoo и др.

Универсальные

простой интерфейс и
удобство для пользователя



■ возможности гибкой
настройки (Custom
Search, Система
поисковых операторов и
функция «расширенного
поиска»)

Специализированные

работают в узкой теме или по
определенному типу контента



■ «Локальные» (учитывают
региональную специфику)



Поисковые операторы

специальные символы, с помощью которых можно уточнить запрос



Оператор	Значение	Примечание
+ (AND) (плюс)	Добавляет слова, которые должны быть в выдаче. В настоящее время избыточен.	Иванов +министр +Беларусь Высоцкий AND Влади
 (OR) (или)	Хотя бы одно из слов	Репортаж OR Новость Драники колдуны
— (минус)	Исключает из запроса «минус-слова» (нежелательные)	БГУ -Бурятия галстук -«песня Сюткина»
«» (точное соответствие)	Точное соответствие запросу (порядок слов, морфология и т. д.)	«Отчет о погрузочных работах за 2018 год» «Бриллиантовая» рука
* (пропущенное слово)	Поиск по цитате с пропущенным словом (словами)	«над пропастью во *»
() (группировка слов)	Для составления сложных запросов	«(трое OR двое) в лодке * и собаки»
loc: (локация)	Поиск по определенному местоположению	университеты loc:Минск

Поисковые операторы

специальные символы, с помощью которых можно уточнить запрос



Оператор	Значение	Примечание
site: (ограничение по сайту)	Результаты поиска только по конкретному веб-ресурсу	Социальные сети site:socmedia.by site:«.by» БГУ (site:by OR site:ru)
inurl: (поиск по URL)	Слово содержится в URL сайта. Оператор allinurl: позволяет найти всю фразу	inurl:бонстики
intitle: (поиск по заголовкам)	Слово содержится в заголовке сайта. Оператор allintitle: позволяет найти всю фразу	intitle:«10 советов журналисту»
intext: (поиск по тексту)	Слово содержится в контенте сайта. Оператор allintext: страница со всеми словами запроса	allintext:курсовые работы по журналистике
source: (источник)	Ищет по конкретному источнику в Google News	Новости Брагина source:БелТА
filetype: (тип файла)	Поиск по определенному типу файлов	Руденко filetype:pdf
cache: (кэшированная версия)	Находит последнюю сохраненную версию сайта	cache:bsu.by

Поисковые операторы

оригинальные символы поисковой системы «Яндекс»



Оператор	Значение	Примечание
! (заданная форма)	Поиск слова в заданной грамматической форме (время, число, падеж)	!выхухоль
& (связанные слова)	Слова встречаются в одном предложении (&) или в пределах документа (&&)	Бах & сыновья
url: (заданный адрес)	Поиск на страницах по заданному адресу	журфак url:bsu.by/*
domain: (заданный домен)	Поиск на страницах в заданном домене	журфак domain:by
host: (заданный хост)	Поиск на страницах на заданном хосте	газеты host:www.blogger.com
mime: (тип файла)	Ищет файлы в указанном формате, которые содержат ключевое слово	диплом mime:doc
lang: (язык)	Поиск на заданном языке	нарыс lang:be

Программируемая поисковая система (Google Custom Search Engine)

Интернет Изображение

Найдено результатов: примерно 1,750,000,000 (за 1.06 сек.)

Melbourne Woman's Dramatic Arrest Sparks Debate On Social Media
https://www.facebook.com/...social-media/950943915405430/?__...
 7 hours ago ... The dramatic arrest of a Melbourne woman who filmed herself being dragged from her car by police officers has sparked debate online.

WSYX ABC 6 - The goal of the social media challenge: for ...
<https://www.facebook.com/WSYXABC6/posts/10156802137319364>
 6 hours ago ... The goal of the social media challenge: for users to guess how much Benadryl they took and to reach a state of hallucination.

How to
<https://www.belta.by>

☐ Сайт	Ярлык	Доступно в Site Restricted JSON API 
☐ belta.by		✓
☐ kp.by		✓
☐ onliner.by		✓



система для поиска файлов
(.doc, .mp3, .pdf и т. д.)



300+ сайтов социальных медиа



250+ сайтов для публикации
видео



поиск по файлообменникам

инструмент позволяет настроить
поиск по определенной выборке
сайтов с гибкой настройкой
результатов выдачи

Nihad A. Hassan, Open Source Intelligence Methods and Tools

Поиск по новостным сюжетам

подбор материалов
на основе интересов

Новости Беларуси (БелТА)
Источник новостей

Подписаться Поделиться

 **Беларусь**

подписка на
отдельные темы и
источники (ленты
новостей)

В новостях

Коронавирус Александр Лукашенко

Владимир Путин Алексей Навальный

Россия Иркутская область

Всемирная организация здравоохранения

Кемеровская область Беларусь

Правила дорожного движения

Google

Глобус Главные новости

Человек Для вас

Звезда Вы подписаны

Лупа Сохраненные запросы

Щит COVID-19

Флаг Россия

Глобус В мире

Местоположение Местные новости

Календарь Бизнес

Лупа Наука и техника

Фильм Развлечения

Велосипед Спорт

Ключ Здоровье

Яндекс

Дата Релевантность

☒ Группировать по сюжетам

☒ Сообщения 2409

☐ Статьи 512

☐ Интервью 122

☐ Видео 25

☐ Фотографии 604

☒ Все

☐ Сегодня

☐ 3 дня

☐ Неделя

☐ За период

группировка схожих
по тематике новостей
в подборки

**ВОЗ зафиксировала рекордное число
заболевших коронавирусом за сутки**

Lenta.ru 00:39
**ВОЗ зафиксировала рекордную
заболеваемость коронавирусом за сутки**
Всемирная организация здравоохранения (ВОЗ) зафиксировала рекордную заболеваемость **коронавирусом** в мире за сутки.

TACC 01:38
**ВОЗ зафиксировала максимальный рост числа
заразившихся коронавирусом в мире за сутки**
ТАСС, 14 сентября. Число подтвержденных случаев инфицирования **коронавирусом** в мире превысило 28,6 млн, увеличившись за 24 часа почти на 308 тыс...

РИА Новости вчера в 23:58
**ВОЗ зафиксировала рекордное число
заражений коронавирусом за сутки**
МОСКВА, 13 сен - РИА Новости. Почти 308 тысяч человек заразились **коронавирусом** в мире за сутки, это рекордный показатель с начала пандемии...

[Ещё 102 сообщения](#)

**7,4 тыс.
источников**

Сервисы упоминаний

позволяют отслеживать обновления по какому-либо ключевому слову – результат регулярно приходит на E-mail



google.ru/alerts



talkwalker.com

Оповещения

Следите за всем новым и интересным в Интернете

По мере появления результатов
Не чаще, чем раз в день
Не чаще, чем раз в неделю

Социальные медиа

Частота отправки	Не чаще, чем раз в день
Источники	Автовыбор
Язык	русский
Страна	Беларусь
Количество	Только лучшие результаты
Доставка	stepanov.pressa@gmail.com

Сохранить Скрыть ▲

☒ Автовыбор
 Новости
 Блоги
 Интернет
 Видео
 Книги
 Обсуждения
 Финансы

RSS-агрегаторы

экспорт содержимого сайта в виде кратких сводок с помощью технологии Reach Site Summary (RSS)



quiterss.org

Программа для компьютера, поддерживает русский язык, дает возможность подписки и управления RSS-лентами. Есть portable-версия



feedly.com

Веб-приложение, бесплатная версия которого позволяет работать с 100 источников информации, которые можно разбить на три категории. Также пользователю доступно три «доски» – раздела, где можно аккумулировать заинтересовавшие материалы

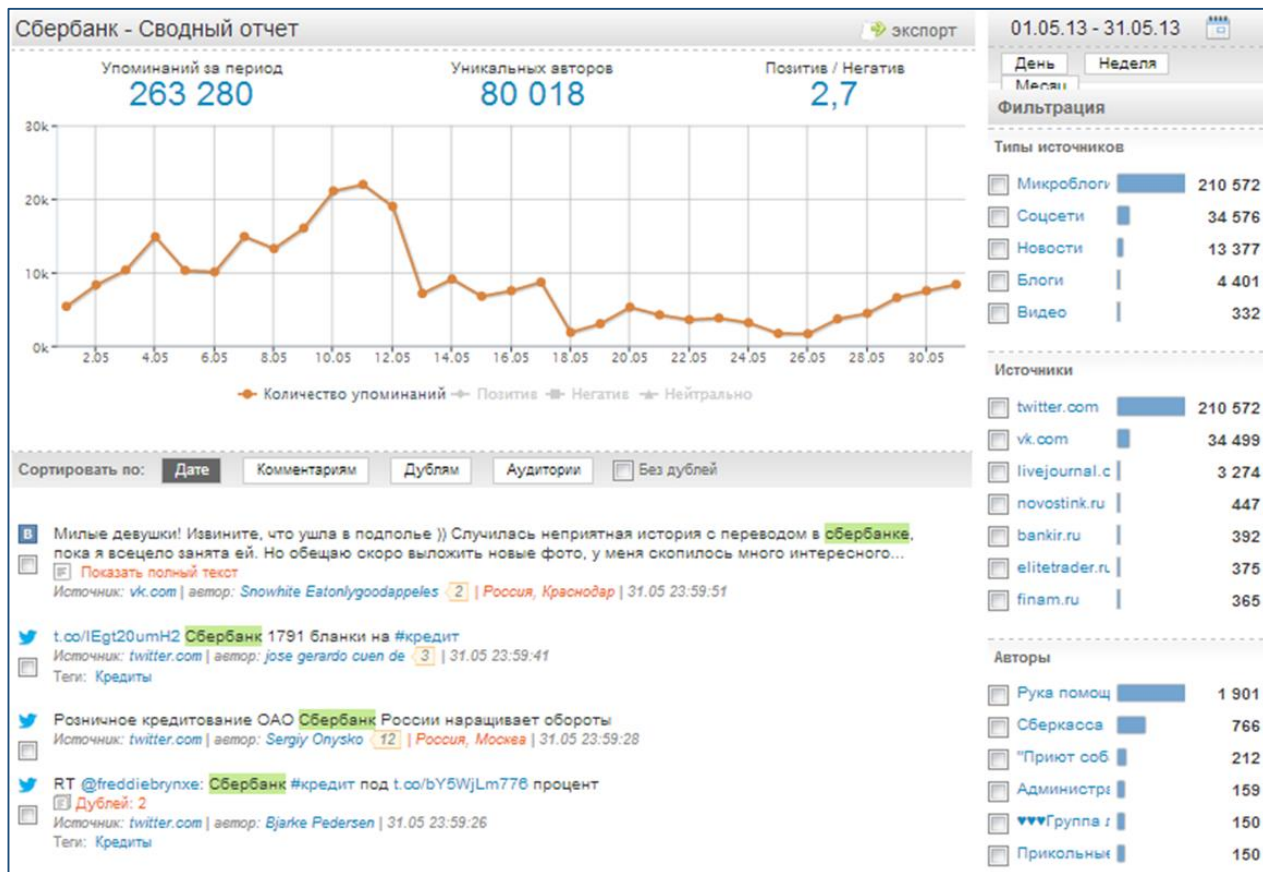


inoreader.com

Сервис для подписки на ленты RSS и для отложенного чтения сайтов, вызвавших интерес, а также мониторинга социальных сетей (в платной версии) До 150 новостных потоков в бесплатном варианте

Сервисы мониторинга социальных медиа

отслеживают социальные сети, блоги, форумы, сайты
отзывов, а также онлайн-СМИ в реальном времени



youscan.io



iqbuzz.pro



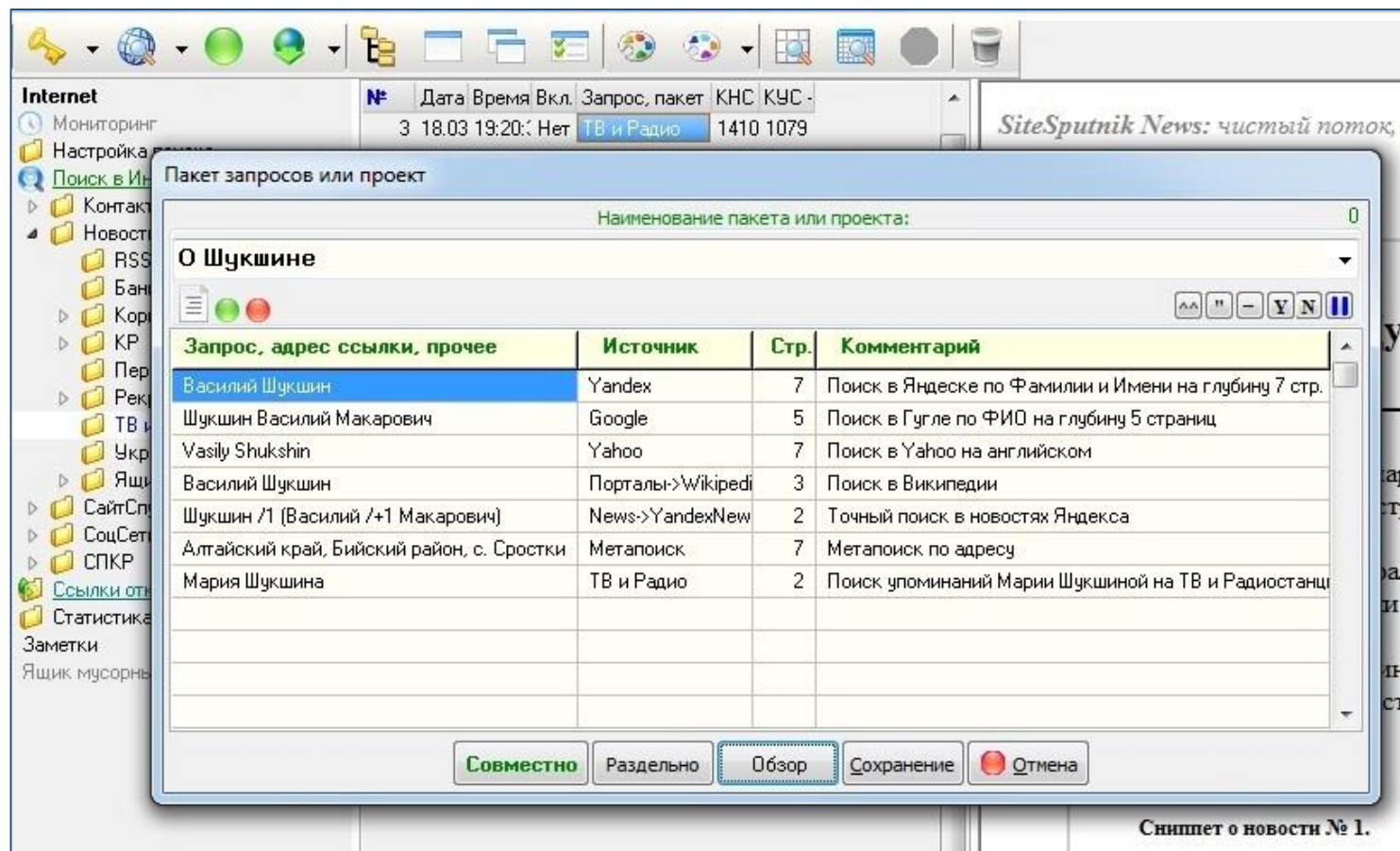
br-analytics.ru



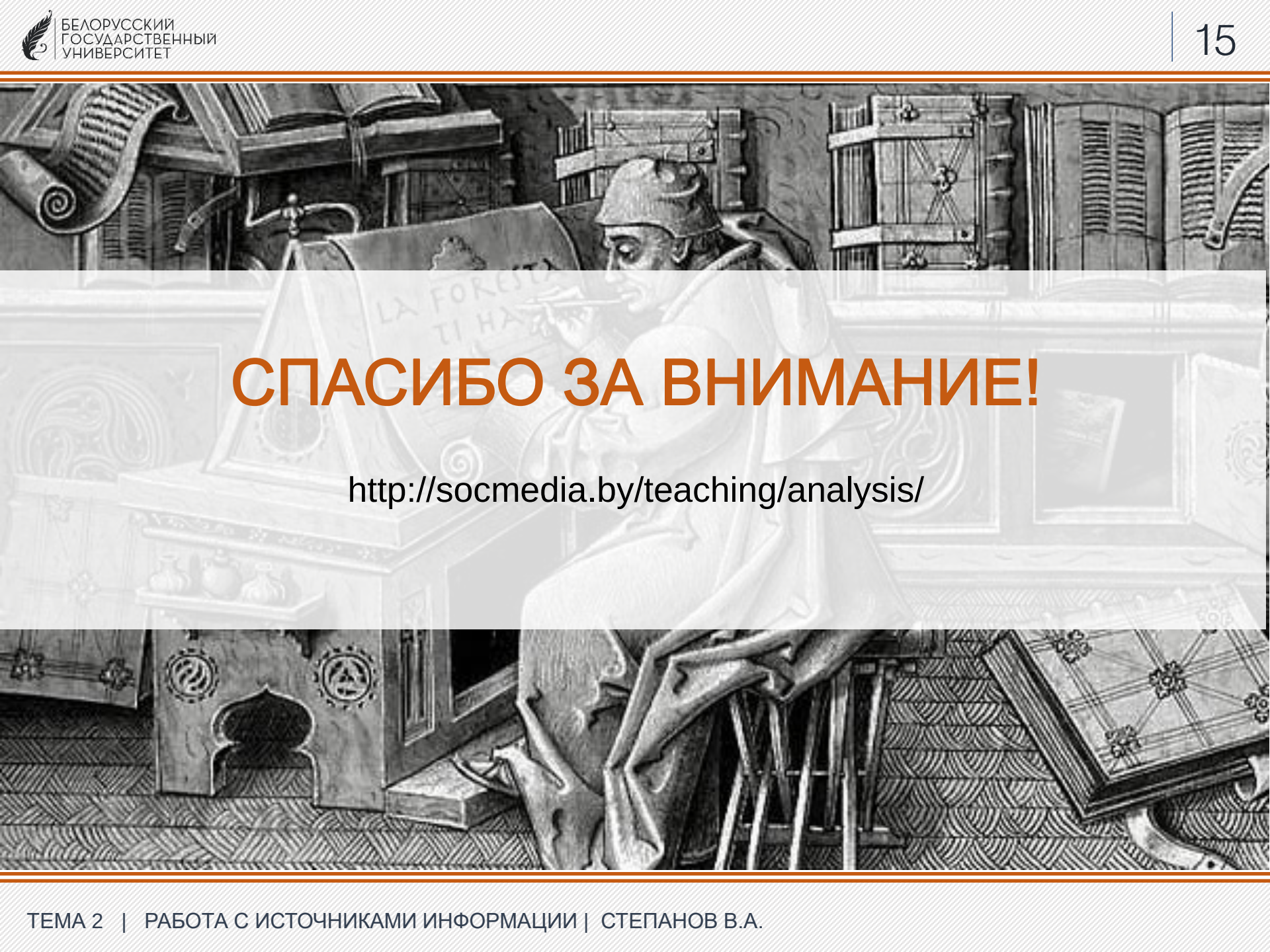
kribrum.ru

Информационно-аналитические системы

автоматизируют процесс поиска сведений, мониторинга сайтов и социальных медиа



Программа
СайтСпутник
(FileForFiles &
SiteSputnik)



СПАСИБО ЗА ВНИМАНИЕ!

<http://socmedia.by/teaching/analysis/>