

ЖИВУЧЕСТЬ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ В СЕТИ ИНТЕРНЕТ

Додонов А.Г., Ландэ Д.В.

*Институт проблем регистрации информации НАН Украины
г. Киев, Украина*

Минск, 21-23 февраля 2013 г.



Понятие живучести информационных объектов в Интернет-среде подразумевает их способность своевременно выполнять свои функции (информирования) в условиях действия дестабилизирующих факторов. Такими факторами могут быть их устранение из информационного пространства, потеря ими свойств актуальности, доступности.



Механизмы обеспечения живучести информационных объектов как систем, являются их неотъемлемой частью, и их эволюция обуславливает эволюцию и механизмов обеспечения живучести. То есть развитие информационных объектов влияет на развитие и механизмов обеспечения живучести.



Задачи, связанные с живучестью информационных объектов в Сети:

- Распространение информационного сюжета
- Устранение информационного сюжета из Интернета
- Защита авторских прав
- Устранение персональных или секретных данных из Интернета
- Конкурентная разведка
- Аналитика (определение тенденций)
- Реклама
- Проведение информационных операций
- ...





Информационные сюжеты как объекты

Информационный сюжет рассматривается как логически связанное множество информационных сообщений, посвященных одной тематике или одному событию. Инфосюжеты являются своеобразными кластерами в информационном пространстве, которые полностью удовлетворяют общему определению систем.

Инфосюжет — это совокупность сильно связанных объектов, обладающая свойствами организации, связности, целостности (определяемой тематикой или событием) и членимостью (на отдельные сообщения). Известно, что системы в общем случае могут разделяться на два класса: целенаправленные и нецеленаправленные. Инфосюжеты как системы могут относиться как к первому, так и ко второму классу. При этом целенаправленные сюжеты могут порождаться в ходе активных рекламных кампаний, в качестве отвлекающих факторов, дезинформирования и других процессов.

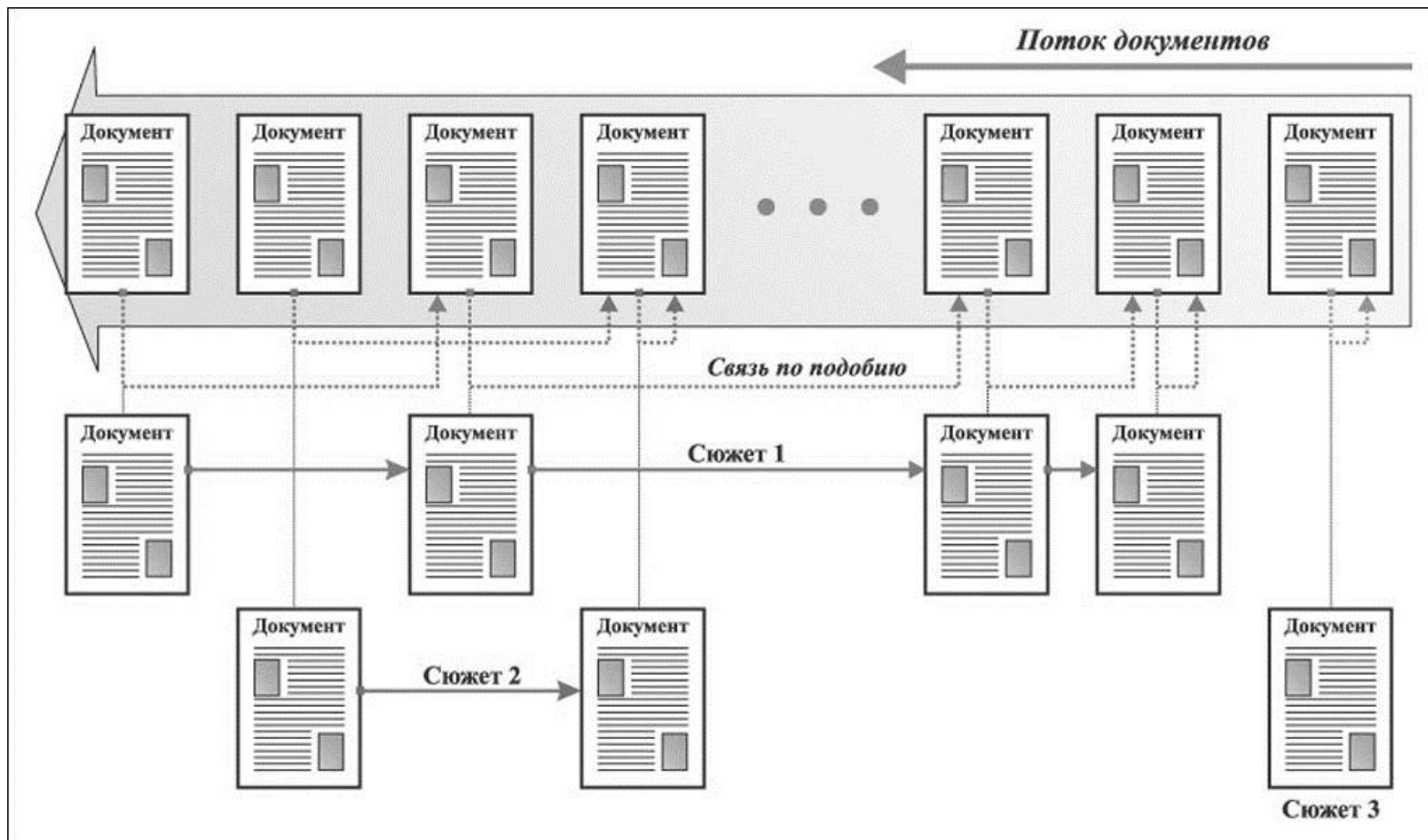


Живучесть информационного сюжета

К потере живучести инфосюжета, например, может привести разрыв связей между его компонентами, например, устранением из информационного пространства наиболее весомых компонент, то есть таких, которые имеют наибольший коэффициент посредничества (betweenness). Этот коэффициент для конкретного узла сети определяется как сумма по всем парам узлов сети соотношений количества кратчайших путей между ними, проходящими через заданный узел, к общему количеству кратчайших путей между ними.



Выявление информационных сюжетов





Роль статистического распределения

Закономерности статистического распределения времени жизни информационных объектов позволяют делать выводы, связанные с их живучестью, а именно учитывать явления самого подобию, нерегулярности потери информации, наличие «тяжелого хвоста» в распределении, которое характеризует чрезвычайно большое количество фактически устаревших информационных объектов и т.п.



Оценки живучести информационных объектов

Живучесть информационного объекта оценивается как вероятность того, что объект будет неповрежденным в течение определенного периода времени t при определенных условиях.

Если информационный объект сохраняется на n носителях информации, то вероятность разрушения этого объекта оценивается как:

$$F_{lost}(t) = \prod_{i=1}^n F_i(t).$$

Частный случай

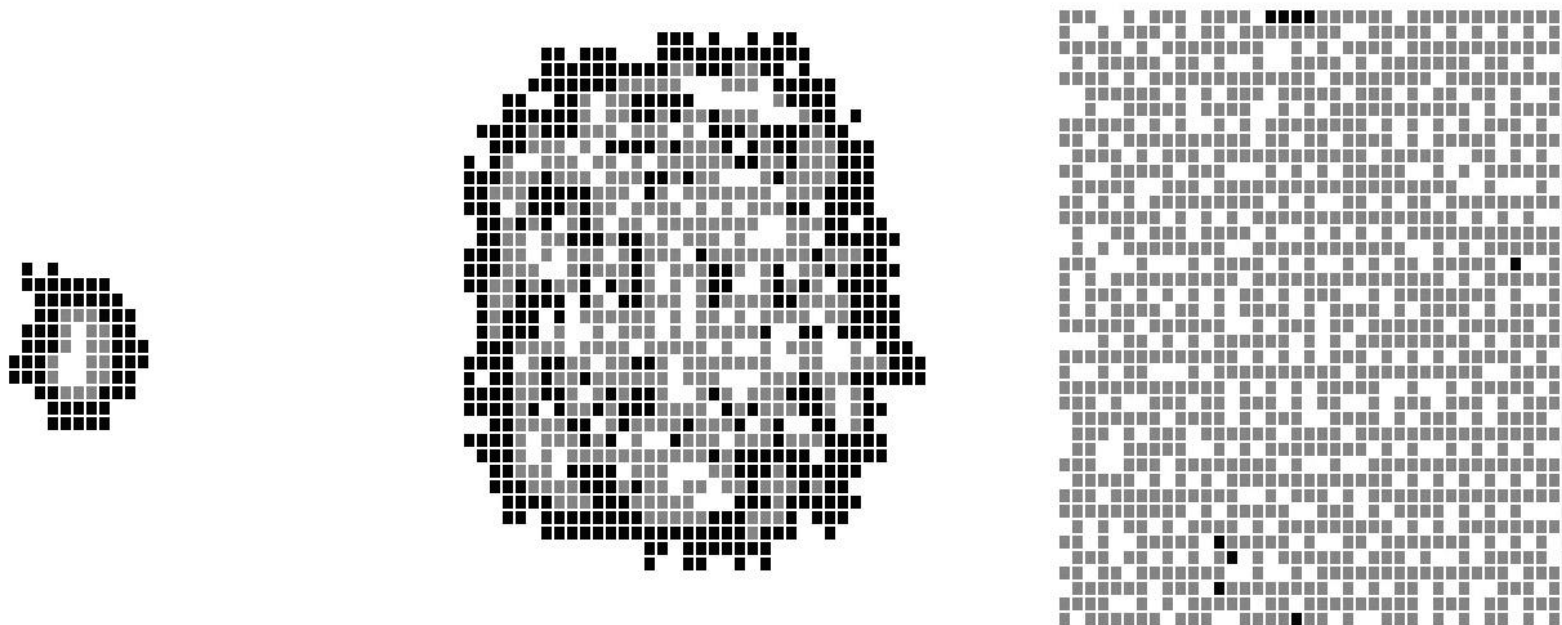
Соответственно, живучесть оценивается как:

$$S_n(t) = 1 - \prod_{i=1}^n F_i(t) = 1 - \prod_{i=1}^n C t^{-\beta} = 1 - C^n t^{-n\beta}$$



МОДЕЛЬ ДИФФУЗИИ ИНФОРМАЦИИ

Жизненный цикл информационных сюжетов может описываться моделью диффузии информации, которая относится к распространению новостей в информационном пространстве. Применяется концепция клеточных автоматов с вероятностными правилами распространения новостей по заданной тематике. Предполагается, что клетка может быть в одном из трех состояний: «свежая новость»; устаревшая, но сохраненная в виде сведений; клетка не имеет информации.

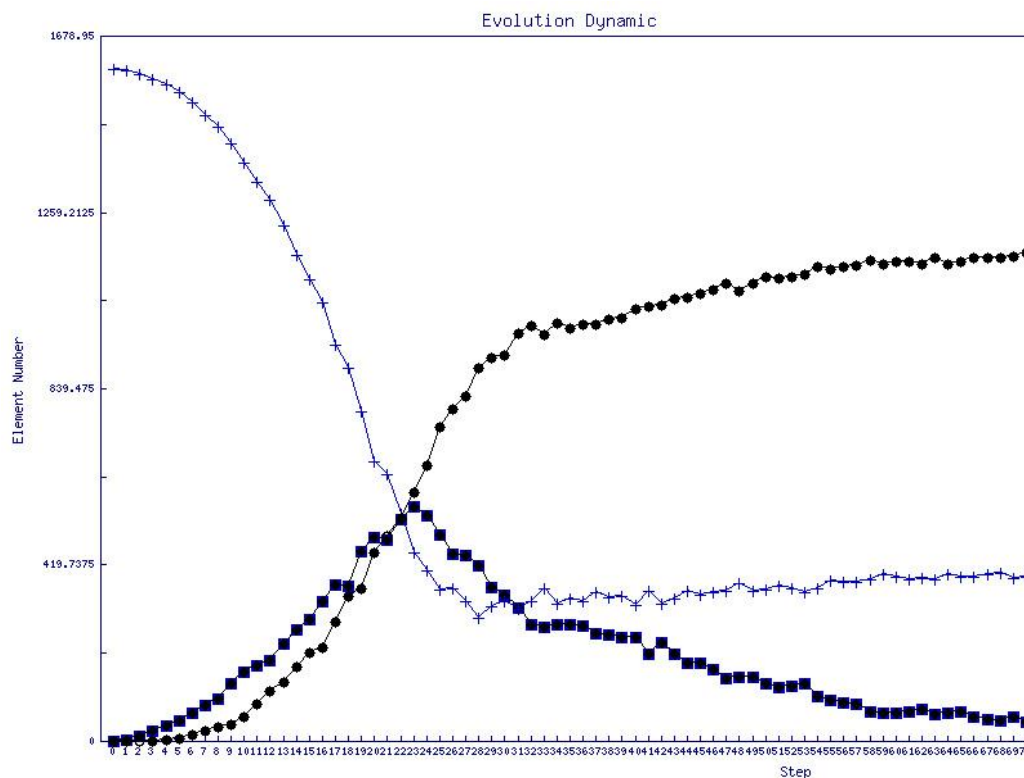




МОДЕЛЬ ДИФфуЗИИ ИНФОРМАЦИИ

Приведенная система клеточных автомат вполне реалистично отражает процесс развития инфосюжета.

Типичные зависимости количества клеток (последовательности количества однотипных клеток), пребывающих в различных состояниях, в зависимости от шагов итерации приведены на рис:

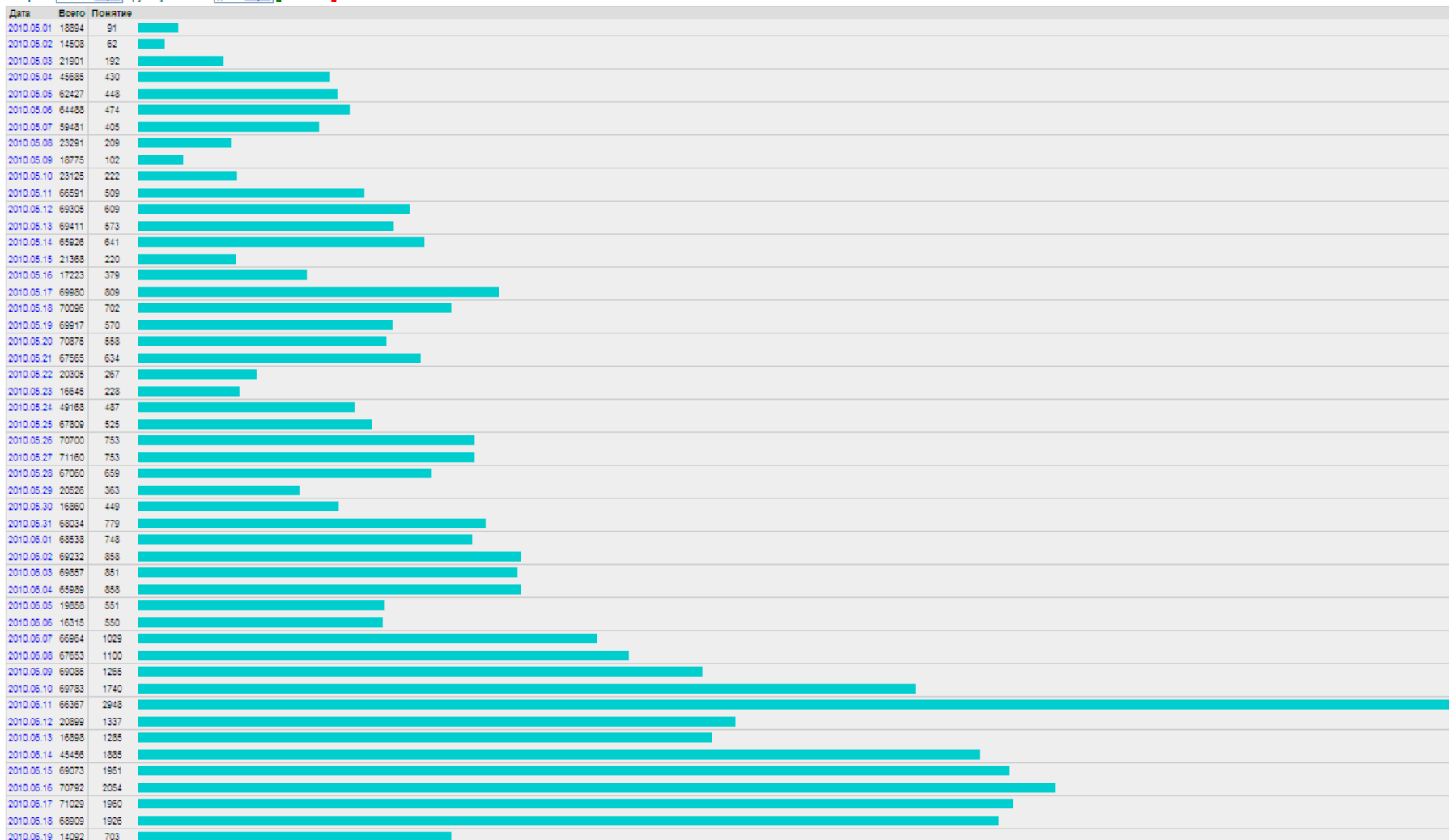




ДИНАМИКА РАЗВИТИЯ СЮЖЕТА (InfoStream)

Понятия в динамике :
+ чемпионат мира футбол

Формат: Группировать по: - позитив - негатив





Существует несколько механизмов, обеспечивающих живучесть информационных объектов в Интернете, рассмотрим лишь наиболее распространенные из них. В реальности эти механизмы применяются не в чистом виде, а как правило, комбинируются.



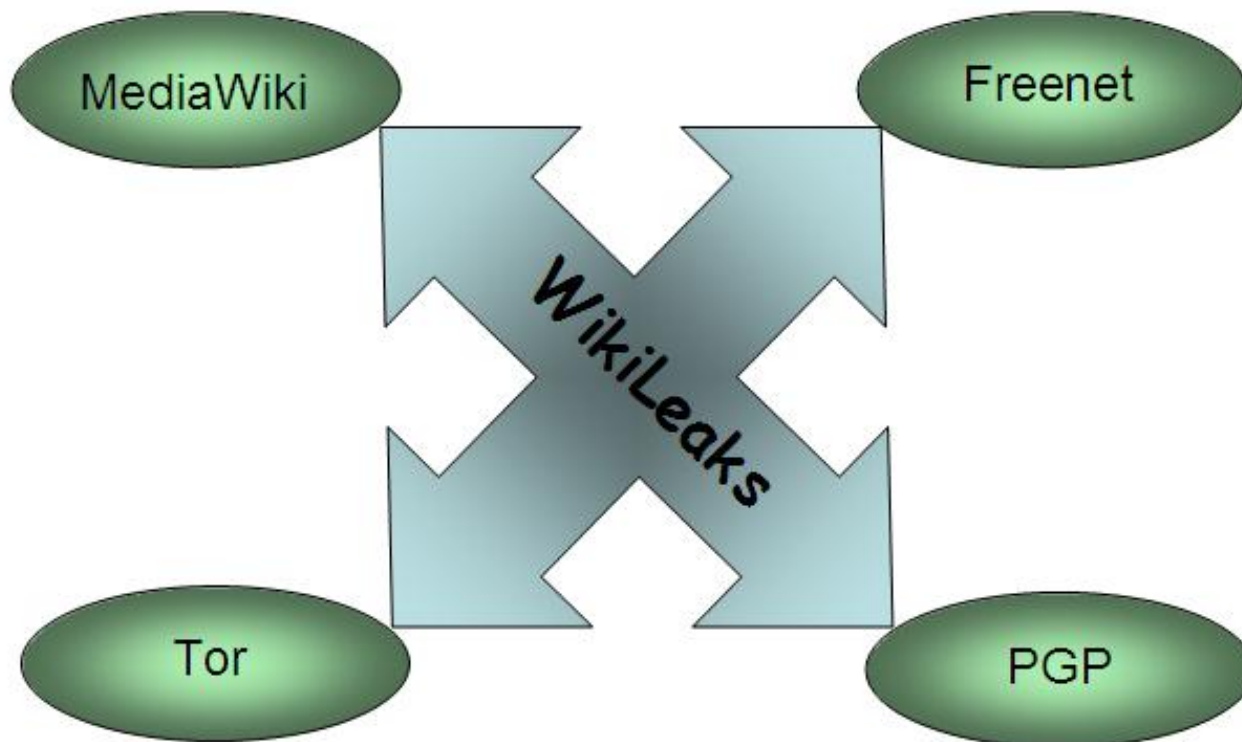
1. Зеркальное копирование данных при размещении их на целевой ресурс. То есть автор размещает информацию, которая копируется хостинг-провайдером на некоторое количество зеркальных серверов. Пример – скандально известная служба WikiLeaks (несколько сотен серверов-зеркал).



Рис. 1 – Копирование информации на зеркальные сервера



4 основных технологии WikiLeaks





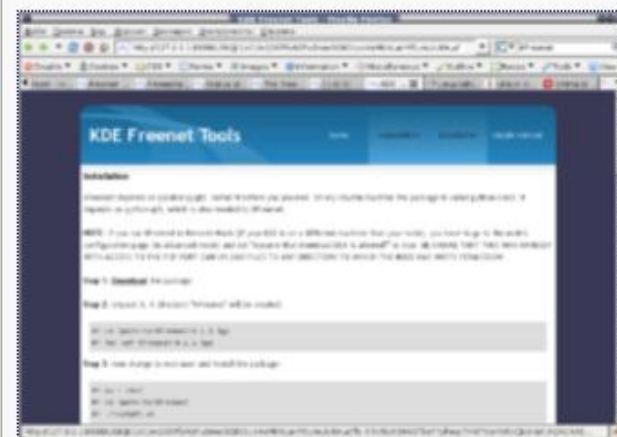
Freenet


Freenet — одноранговая сеть, предназначенная для децентрализованного распределённого хранения данных без возможности их цензуры, созданная с целью предоставить пользователям электронную свободу слова путём обеспечения их строгой анонимности. Freenet работает на основе объединения в общий фонд (пулинга) предоставленной пользователями (членами сети) своей полосы пропускания и дискового пространства своих компьютеров для публикации или получения из Freenet разного рода информации. Freenet использует разновидность маршрутизации по ключам, похожей на распределённую хеш-таблицу, для определения местонахождения пользовательских данных^[1].

Freenet находится в стадии разработки с 2000 года. Хотя версия 1.0 ещё не выпущена, текущие версии достаточно стабильны для практического использования.

По мнению многих, Freenet фундаментально отличается от других пиринговых сетей: эта сеть все ещё несколько сложнее в использовании и значительно медленней. Однако после включения и некоторого времени работы в этой сети её узлы становятся более быстрыми и начинают более эффективно извлекать из сети данные.

Freenet



Тип	анонимная одноранговая сеть
Разработчик	Команда разработчиков Freenet 



2. Перепечатка информации (републикации, «копипаст») на другие сайты (рис. 2).



Рис. 2 – Републикация информации



InfoStream
Online



Формат: Обычный Группировать по: дням - позитив - негатив

Динамика



Дата	Всего	Понятие
2012.09.01	26046	26046
2012.09.02	21632	21632
2012.09.03	74877	74877
2012.09.04	77873	77873
2012.09.05	79034	79034
2012.09.06	79261	79263
2012.09.07	75595	75593
2012.09.08	27275	27271
2012.09.09	23204	23208
2012.09.10	76272	76272
2012.09.11	78328	78328
2012.09.12	79859	79859
2012.09.13	78842	78842
2012.09.14	75546	75546
2012.09.15	28155	28155



Соотношение оригинальной информации и общего объема информации, сканируемой системой InfoStream за первые четыре месяца 2012 г. по дням приведено на рис. 3. Следует отметить, что наиболее важная и интересная информация перепечатывается сотни раз, в то время как неактуальная практически не дублируется.

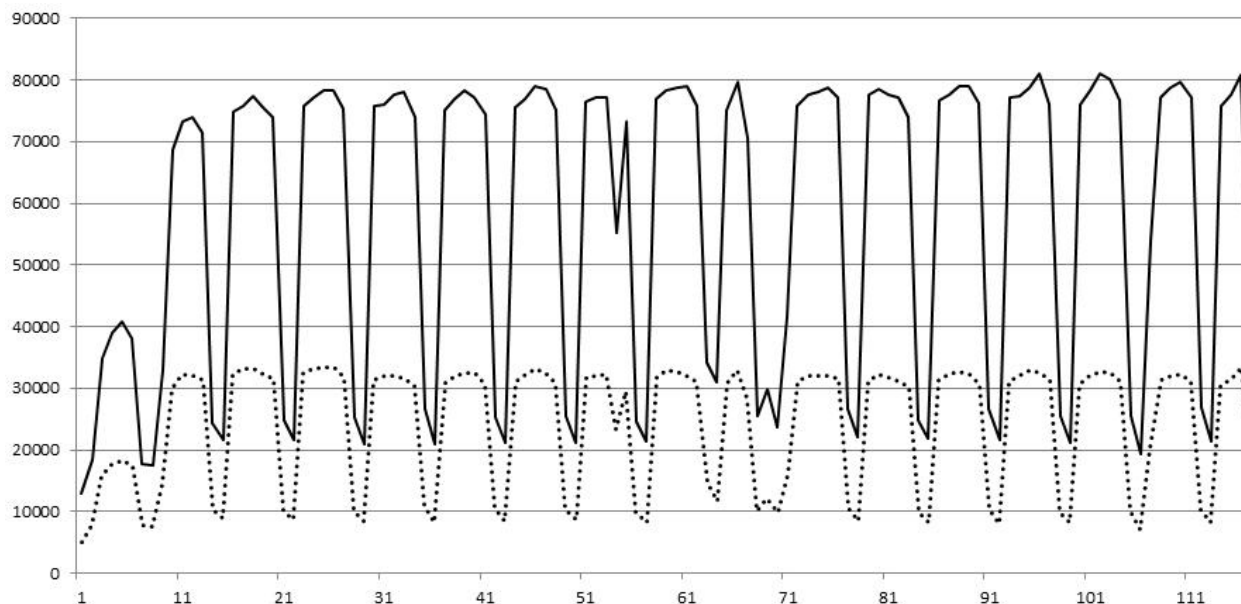


Рис. 3 – Соотношение оригинальной информации (пунктирная линия) и общего объема информации (сплошная линия)



3. Размещенная однажды информация навсегда попадает в архивные службы Интернета типа Internet Archive (рис. 4).

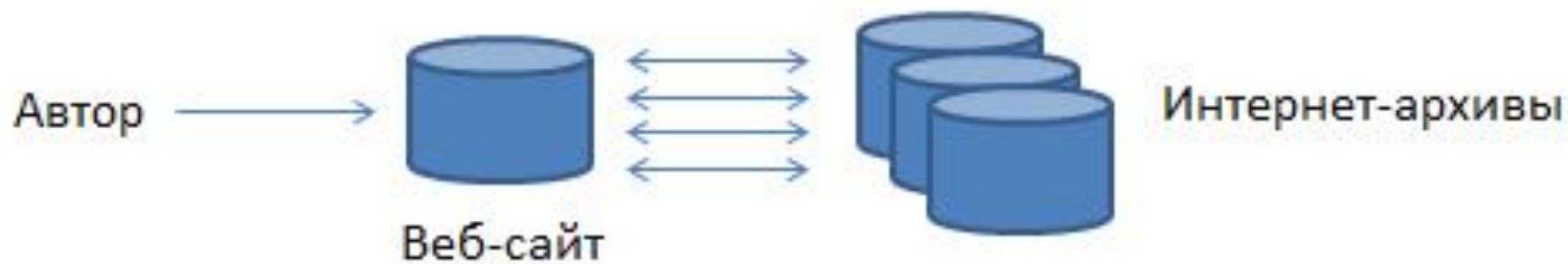


Рис. 4 – Сохранение в Интернет-архивах



[Web](#) [Video](#) [Texts](#) [Audio](#) [Projects](#) [About](#) [Account](#) [TVNews](#) [OpenLibrary](#)

Universal access
to all knowledge

[Home](#) [Donate](#) | [Forums](#) | [FAQs](#) | [Contributions](#) | [Terms, Privacy, & Copyright](#) | [Contact](#) | [Volunteer Positions](#) | [Jobs](#) | [Bios](#)

Search:

FAQs



Advanced Search

Anonymous User (login or join us)

Upload

Frequently Asked Questions

[The Wayback Machine](#) | [The Internet Archive](#) | [Audio](#) | [Live Music Archive](#) | [Texts and Books](#) | [Movies](#) | [Downloading Content](#) | [Virtual Library Cards \(AKA Accounts\)](#) | [FreeCache](#) | [DocuComp](#) | [Prelinger Movies](#) | [Search Tips](#) | [Uploading Content](#) | [Archive BitTorrents](#) | [Forums](#) | [SFLan](#) | [Report Item](#) | [Rights](#) | [Equipment](#) | [Archive-It](#)

Questions

What is the Wayback Machine?
How can I get my site included
in the Wayback Machine?

[How can I have my site's
pages excluded from the
Wayback Machine?](#)

The Wayback Machine

What is the Wayback Machine? How can I get my site included in the Wayback Machine?

The Internet Archive [Wayback Machine](#) is a service that allows people to visit archived versions of Web sites. Visitors to the Wayback Machine can type in a URL, select a date range, and then begin surfing on an archived version of the Web. Imagine surfing circa 1999 and looking at all the Y2K hype, or revisiting an older version of your favorite Web site. The Internet Archive Wayback Machine can make all of this possible.

How can I get my site included in the Wayback Machine?



ASK A LIBRARIAN

DIGITAL COLLECTIONS

LIBRARY CATALOGS

Search

GO

[The Library of Congress](#) > [News Releases](#) > [Twitter Archive to Library of Congress](#)

NEWS RELEASES



Search this site

GO

- [News Releases Home](#)
- [Events: Talks, Films, Concerts](#)
- [Library of Congress Blog](#)

News from the Library of Congress

Contact: Matt Raymond, Library of Congress, (202) 707-2905, mray@loc.gov
Contact: Greg Pass, Twitter, gregpass@twitter.com

April 15, 2010

Twitter Donates Entire Tweet Archive to Library of Congress

Twitter is donating its digital archive of public tweets to the Library of Congress. Twitter is a leading social networking service that enables users to send and receive tweets, which consist of web messages of up to 140 characters.

[Print](#) [Subscribe](#) [Share/Save](#)



4. Информация индексируется глобальными поисковыми системами типа Google, Яндекс и остается в их кэш-памяти, откуда она доступна пользователям (рис. 5).

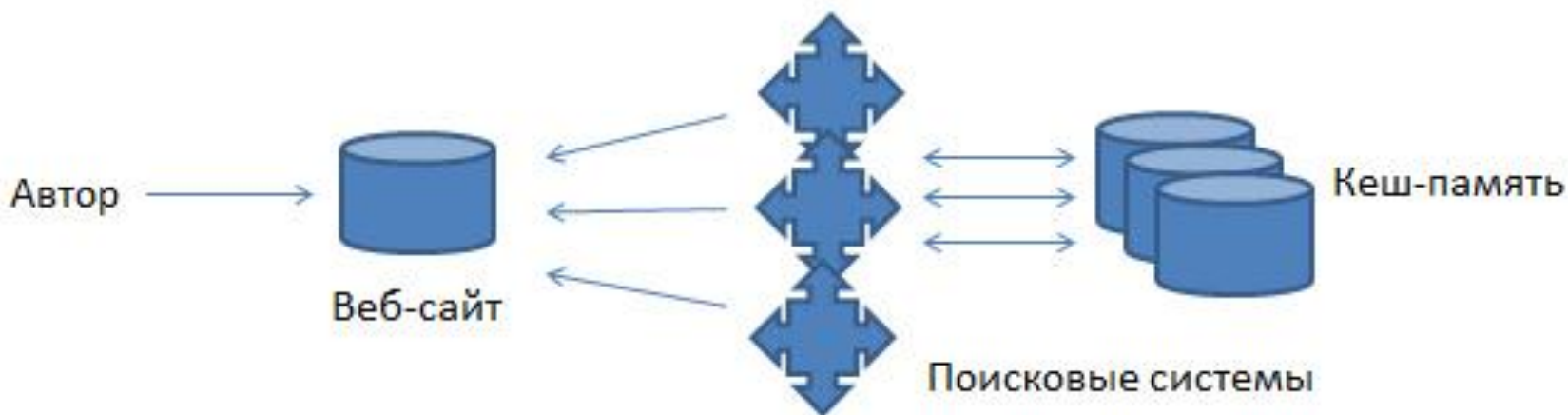


Рис. 5 – Сохранение информации в кэш-памяти поисковых систем



Сообщений на страницу 25

[ровано](#), [Опубликовано](#) 1 – 25 из 280 [Старые](#) [Самые старые](#) »

Комментарии: 2	13.12.11	автор(ы): Вадим	Удалить
Комментарии: 5	10.12.11	автор(ы): Вадим	Удалить
Комментарии: 6	09.12.11	автор(ы): Вадим	Удалить
Комментарии: 7	08.12.11	автор(ы): Вадим	Удалить
Комментарии: 2	08.12.11	автор(ы): Вадим	Удалить

Google

Яндекс

[Вебмастер](#) [Метрика](#) [Виджеты](#) [Рекламная сеть](#) [Директ](#) [Поиск для сайта](#) [API](#) [еще](#)

[вебмастер](#)

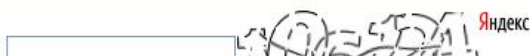
[Мои сайты](#) [Додати сайт](#)

Видалити URL

Якщо сторінка сайту більше не існує або ви не хочете, щоб вона з'являлась у результатах пошуку, то ви можете прискорити процес видалення цієї сторінки з бази Яндекса, зазначивши її URL у наведеній формі. Для видалення необхідно, щоб під час запиту зазначеного URL сервер повертав код помилки 404 або індексування сторінки було заборонено у файлі robots.txt, або метатегом noindex.

i Стандартний протокол http зазначати не обов'язково. Якщо ж доступ до сайту обмежений і здійснюється за [https](#), протокол потрібно зазначити явно. Наприклад, <https://mail.yandex.ru>

URL



Яндекс



5. Информация с веб-сайта может сохраняться на локальных компьютерах конечных пользователей, которые получили к ней доступ либо непосредственно, либо через интеграторов информации (рис. 6).



Рис. 6 – Сохранение информации на компьютерах пользователей



Удаление информации с веб-ресурса – не может гарантировать ее исчезновение из Интернета. Остаются не только «цифровые следы» и «цифровые тени».

Выражение «цифровые следы» («digital footprint») относятся к той информации, которая оставляется самим пользователем при работе в Сети.

Информация о пользователе, создаваемая без его участия, получила название «цифровой тени» («digital shadow»).





Интернет-чистильщик - удаляет все следы
вашей деятельности в Интернете.
Пример: “suicide machine”

[About](#) [FAQ](#) [Take a Tour](#) [Review](#) [Download](#) [Contact](#)

 **web^{2.0} suicide machine**

*Wanna meet your real
neighbours again?*



Select your
Social Network:



Please read the C&D letter from Facebook [here!!!](#)



Еще один пример - Интернет-чистильщик ВебШериф по заказу правообладателей находит Интернет-ресурсы, распространяющие ворованный контент. В основном это фильмы, игры, музыка и клипы.

ВЕБШЕРИФ
ОТРЯД ПО БОРЬБЕ С ПИРАТСТВОМ

+7(495)739-2550
Звоните нам по будням с 10:00 до 18:00 или
пишите в любое время на sheriff@websheriff.ru

Новости | Контент под защитой | Технологии | Стоимость | Обратная связь

Вход для клиентов
логин

войти

3 прошедшие **А**сутки

УДАЛЕНО	ОТСЛЕЖИВАЕТСЯ	ОТПРАВЛЕНО
001332	03602	00061
Удалено нелегальных файлов	Сайтов отслеживается системой	Отправлено жалоб на нелегальный контент

Статья 1250: Защита интеллектуальных прав

TM

Важная информация, попав в Интернет, остается там практически навсегда, и как показывает практика, рассчитывать на ее легкое удаление или изменение не приходится. Лучшим методом оказывается вытеснение нежелательной информации новыми сюжетами, проведение специальных мероприятий по содержательному исправлению ошибок.

СПАСИБО ЗА ВНИМАНИЕ!

Дмитрий ЛАНДЭ
dwlande@gmail.com