

[\[\[Википедия\]\]](#)

Проект Качество

Проект Связность

# Связность статей нашего раздела

**Участник: Mashiah Davidson**

2007 год

# План доклада

- ▶ **Что такое связность?**
- ▶ **Страницы-сироты, тупиковые и изолированные статьи**
- ▶ **Инструменты для анализа связности**
- ▶ **Сколько изолированных кусков в Википедии?**
- ▶ **Как повышают связность**
- ▶ **Открытые вопросы для обсуждения на круглом столе**

# Что такое связность?

- ▶ **Связи:**
  - ▶ **[[Гиперссылки]]** — направленные связи между статьями
  - ▶ ↪ Страницы перенаправления
- ▶ **Связи улучшают навигацию и повышают верифицируемость информации**
- ▶ **Все статьи должны быть доступны при переходе по ссылкам с Главной страницы**
- ▶ **Страница не считается статьёй, если она не содержит ссылок на другие страницы**
- ▶ **Двойные и ошибочные перенаправления спрямляются**



# Страницы-сироты

- ▶ **Статьи основного пространства имён, на которые нельзя перейти по гиперссылке при прочтении других статей**



На эту статью не ссылаются другие статьи Википедии.

Пожалуйста, воспользуйтесь [поиском](#) и установите ссылки в соответствии с [принятыми рекомендациями](#).

- ▶ Страницы перенаправления не бывают сиротами
- ▶ Страницы значений не бывают сиротами, ссылки из них не учитываются

---

*Список статей со сходными заголовками.*



*Если Вы попали сюда из другой статьи Википедии, пожалуйста, вернитесь и исправьте ссылку непосредственно на верную статью.*

---

- ▶ **Ссылки** из общехронологических и тематических хронологических статей не учитываются



## 8 енгабря в хронологии

Материал из Википедии — свободной энциклопедии

### Родились 8 енгабря

- [Вася Пупкин](#)

# Тупиковые статьи

- ▶ Статьи основного пространства имён, не ссылающиеся на другие статьи
- ▶ Ссылки (прямые или через перенаправление) на страницы значений не учитываются
- ▶ Ссылки в другие пространства имён не учитываются

```
{{rq|wikify|style|translate|sources|stub|coord}}
```

Для улучшения статьи желательно:

- [Викифицировать](#) статью
- Привести стиль [в соответствие](#) с нормами современного русского языка
- Перевести текст [Википедия:Викификация](#) на русский
- Найти и указать [авторитетные источники](#), подтверждающие написанное
- Дополнить статью хотя бы до размеров [заготовки](#)
- Добавить [географические координаты](#)



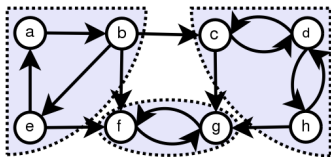
- ▶ **Красные ссылки** не учитываются

Очепятка в ссылке.

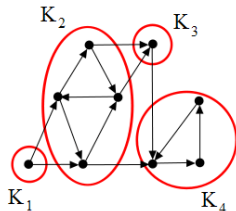
Очепятка

# Изолированные статьи

- ▶ Кластеры — сильно связанные компоненты графа ссылок



- ▶ Цепочки из кластеров



# Связность в движке MediaWiki

- ▶ **[[Служебная:Lonelypages]]** — автоматически формируемый список страниц-сирот (около 2,5 тысяч)
- ▶ **[[Служебная:Deadendpages]]** — автоматически формируемый список тупиковых страниц (около 100)
- ▶ Ссылки из других пространств имён, со страниц значений, из «сиротливых» перенаправлений, из хронологических статей учитываются наряду со ссылками из тематически близких статей.
- ▶ Неудобны для совместной работы, поскольку обновляются один раз в день.

- ▶  **ссылки сюда**

# Проект:Связность

## ▶ «Качество:Связность»



Участник проекта  
«Связность»

**15 участников**

- ▶ **поиск по Википедии**
- ▶ **поиск синонимов и создание страниц перенаправления**
- ▶ **написание и пополнение статей, в том числе перевод**
- ▶ **викификация статей**
- ▶ **уточнение ссылок, ведущих на страницы значений**
- ▶ **вынесение статей на улучшение, объединение, удаление**
- ▶ **улучшение инструментов для анализа и повышения связности**



Приглашаю участников Википедии в проект «Связность». Это ТРУДНО, ИНТЕРЕСНО и НАДОЛГО => [Mashiah](#) 23:21, 5 октября 2007 (UTC)



# Деятельность бота «Голем»

- ▶ Околоежедневный сбор списков тупиковых и изолированных статей
- ▶ Установка и снятие шаблонов {{сирота}}, {{тупиковая}}, {{изолированная статья}}
- ▶ Список двойных перенаправлений, более полный чем [[Служебная:DoubleRedirects]]
- ▶ Список перенаправлений, содержащих более одной ссылки
- ▶ Инструмент для анализа изолированных категорий:  
<http://hemlock.knams.wikimedia.org/~mashiah/cgi-bin/category.sh>

# Сколько связности в Википедии

- ▶ 14,5 тысяч изолированных статей из 195 тысяч (7,5%)
- ▶ если не считать списки статьями, будет около 16,5 тысяч (8,5%)
- ▶ 11,9 тысяч страниц-сирот (6%), в августе было 12,3
- ▶ до запуска бота число сирот росло так же, как число страниц значений, которых теперь 14,5 тысяч
- ▶ 1 000 тупиковых статей (0,5%)
- ▶ 900 изолированных статей типа \_1\_1
- ▶ 700 изолированных статей типа \_2
- ▶ 165 изолированных статей типа \_3
- ▶ 127 изолированных статей типа \_1\_1\_1
- ▶ 62 изолированные статьи типа \_1\_2
- ▶ 54 изолированные статьи типа \_2\_1
- ▶ 51 изолированная статья типа \_1\_51
- ▶ 36 изолированных статей типа \_4
- ▶ ограничение на учёт кластеров размером больше 10
- ▶ размер основного кластера: 176174

# Сколько связности в категориях

- ▶ **Больше всего изолированных статей**
  - ▶ Категория:Персоналии по алфавиту — 6,5 тысяч
  - ▶ Категория:Материалы ЭСБЕ — 3,65 тысяч
  - ▶ Категория:Незавершённые статьи о людях — 2,75 тысяч
  - ▶ Категория:Населённые пункты по алфавиту — 372
  - ▶ Категория:Ныне живущие — 368
  - ▶ Категория:Незавершённые статьи об учёных — 352
  - ▶ Категория:Мультфильмы по алфавиту — 319
  - ▶ Категория:Незавершённые статьи — 303
- ▶ **Среди служебных категорий**
  - ▶ Категория:Википедия:Статьи без интервик — 1 450
  - ▶ Категория:Википедия:Статьи к викификации — 1 063
  - ▶ Категория:Википедия:Статьи без ссылок на источники — 601
  - ▶ Категория:Википедия:Статьи без иллюстраций — 502
  - ▶ Категория:Википедия:Тупиковые статьи — 405
  - ▶ Категория:Википедия:Очень короткие статьи — 277
- ▶ **Актуально**
  - ▶ Категория:Материалы БСЭ — 27, 15 сентября было 77

# Как повышают связность

## ▶ **Викификация статей (в том числе тупиковых)**

- ▶ Связная речь и связный текст
- ▶ Разбиение статей на разделы
- ▶ Ссылки на контекст, окружение и частности
- ▶ Межъязыковые ссылки
- ▶ Категории для названия статьи
- ▶ Оценка законченности статьи и её викификации

## ▶ **Установка ссылок на изолированные статьи**

- ▶ Ссылки сюда: а нет ли там страниц значений?
- ▶ [[Википедия:Исправление ссылок на неоднозначности]]
- ▶ Начните поиск с шаблона {{сирота}}
- ▶ Викификация найденных статей
- ▶ Не ограничивайтесь одной ссылкой
- ▶ Поиск синонимов и создание перенаправлений
- ▶ Создание новых статей-эссе

## ▶ **Улучшение инструментов**

- ▶ Сделайте запрос на инструмент

## **Общение с новыми участниками**

# Голем изнутри

- ▶ Тул-сервер в Нидерландах
  - ▶ yarrow: sql\_s3: БД для русского языка → isolated.sql →
  - ▶ hemlock: Клиент Linux → isolated.sh → файлы →
- ▶ Клиент Windows → AWB → Википедия
- ▶ Исходный код:
  - ▶ <http://fisheye.ts.wikimedia.org/browse/mashiah/isolated>
- ▶ Планы
  - ▶ yarrow: sql\_s3: БД для русского языка → isolated.sql →
  - ▶ hemlock: Linux консоль → isolated.sh → ???..pl → Википедия

# Вопросы для круглого стола

## ▶ Шаблоны

- ▶ Видимый шаблон или скрытый?
- ▶ Начало статьи или конец?
- ▶ Общий служебный шаблон `{{rq}}`
- ▶ Кто умеет рисовать?

## ▶ Учёт

- ▶ Именованние служебных категорий для изолированных статей
- ▶ Внедрение в других языковых разделах

## ▶ Алгоритмы

- ▶ Задержка между правками участников и бота
- ▶ Алгоритм кластеризации направленного графа
- ▶ Учёт списков для координации работ
- ▶ Неполнота списка двойных перенаправлений

## ▶ Правила

- ▶ [\[\[Википедия:Статья\]\]](#)
- ▶ [\[\[Википедия:Ссылка\]\]](#)
- ▶ [\[\[Википедия:Изолированные статьи\]\]](#)
- ▶ [\[\[Википедия:Страницы-сироты\]\]](#)
- ▶ [\[\[Википедия:Тупиковые статьи\]\]](#)