

# Поиск информации (в Интернете)

# Предупреждения

1. Критерий качества тех алгоритмов поиска, которыми Вы пользуетесь – результат и оценка Ваших (исследовательских) проектов. Если Вы и окружающие этими проектами довольны, - значит, можно работать без дополнительных ухищрений, как привыкли.
  - Пока не изменился уровень амбиций и Интернет.
2. Даже успешный поиск достигается разной ценой. Хорошая формулировка запроса – меньше времени на ручной отбор результатов.
3. Я тоже пользуюсь 10 % того, что собираюсь упомянуть. Но знание об этих 90 % - явно полезно.
4. Я буду признателен за любые дополнения этого списка другими полезными ресурсами, ссылками, советами (и презентациями?).

# План

1. Знать, что хочешь найти
2. Какие слова вбивать
3. Куда вбивать ключевые слова
4. Если поиск не удался
5. Содержательная экспертиза результатов поиска
6. То, что находится параллельно поиску. Типы дополнительных ресурсов, обнаруживающиеся через Интернет
7. Вместо резюме: поиск и серфинг

# 1. Знать, что хочешь найти.

## Специфика Интернета

- ▶ Информации очень много и она предоставляется (поисковыми машинами и т.п.) очень быстро и очень удобным образом.
- ▶ (основное различие с оффлайн-миром) Информация – какая угодно. Непроверенная, ангажированная, случайная, черновая. Способная ввести в заблуждение относительно квалификации или (отсутствия) интересов автора или ресурса. Редко указываются предшествующие источники информации.

Пока Интернет опережает социальный контроль.

Следовательно, экспертизу приходится осуществлять пользователю.

- ▶ Информация не обязательно упорядочена и описана в соответствии с логикой и потребностями Вашего исследовательского проекта. Различие логик ранжирования разных поисковых машин, да и составления каталогов.
- ▶ Многоэтажность» обычного поиска
  - "Найти санаторий на Финском заливе, принимающий семьи с детьми двух лет, с хорошими отзывами, низкими ценами и свободными за месяц местами на первую декаду июля 2007 года"

Рекомендация: уже при придумывании поиска четко представлять себе характер разыскиваемого контента:

- ▶ Общий обзор проблемы
- ▶ Полемика разных точек зрения
- ▶ Детальные экскурсы в проблему, способные предоставить аргументированное обобщение данных для Вашего исследования.

## Различать поиск «для идей» и поиск «для ДОВОДОВ»

Рыскаете ли Вы в поисках новых идей и постановок вопросов, или в поисках данных, которые подтвердили или опровергли бы концепцию?

**Допустим, для выполнения задания по интеллектуальному ландшафту Москвы (на 30.09) надо ясно понимать, что в первую очередь надо найти новостные ресурсы, извещающие о ближайших и состоявшихся научных мероприятиях**

# Знать, что хочешь найти. «Модель знаний»

Допустим, для выполнения задания по интеллектуальному ландшафту Москвы (на 30.09) надо ясно понимать, что центральной категорией последующего анализа (когда Вы будете вырабатывать свои предложения о семинарах) будут не «новости» и не «мероприятия», а **«центры»**: какие люди, в каких институциях и благодаря какой мотивации занимаются Возрождением или историей радио.

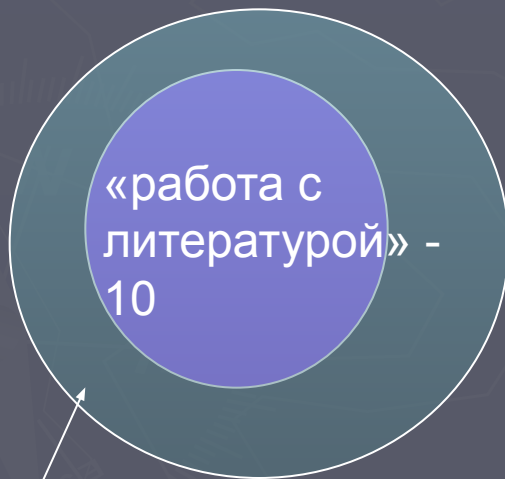
- ▶ От этого уже будут зависеть столбцы Вашей «таблицы»: по каким параметрам описывается мероприятие («место, время», «организаторы», «следы деятельности в Интернете», «с какими дисциплинами пересекается», «финансирование») и т.п.
- ▶ На первичный поиск (новостных лент о семинарах), строго говоря, не влияет ни то, ни другое. Но на просмотр результатов – уже да, и лучше осознавать, по ячейкам какого шкафа Вы собираетесь раскладывать полученную информацию.

## 2. Какие слова вбивать

Содержание  
поискового запроса.  
«Концентрический»  
принцип

Набор ключевых слов  
(«лучевая схема» abstract'a)  
нуждается в переводе

«Радиусы»  
ключевых слов



«научное сообщество»,  
«коммуникация», «научные  
революции» - 40

1. Некомпетентность (во всем) библиографов и рубрикаторов каталогов в Интернете.
2. Устаревание исследований и авторов (то, что в 1980-х-2000-х называется "gender", раньше описывалось как "sexual").

Советы: удобны персоналии и термины. Существительные в именительном падеже. Иностранные языки.

Ключевые слова из поля источника и поля методологии. Обязательность слов, которых нет в источнике – опознание своей научной традиции.

## 2. Какие слова вбивать

- ▶ Допустим, ищете текст статьи «О задаче историка» Гумбольдта
  - Что вобьете в поисковую строку?
  - «Скачать» «О задаче историка» «Во всем происходящем действует не воспринимаемая непосредственно идея и что познана эта идея может быть только в пределах самих событий»
- ▶ Хотите узнать, как делается сыр (Яндекс)
  - *(история, технология, изготовление) /+1 (сыра, творога)*
- ▶ Хотите продать “deformed bars” (Гугл)
  - В поле «с точной фразой» = “deformed bars”, «с любыми словами» = “buy CNF CFR C&F purchase”, «без слов» = “sale”, «время» = «в течение последних трех месяцев». Можно пытаться ограничивать регион, язык, формат.



# Синтаксис поискового запроса

- ▶ Расширенный поиск (цель – сузить зону поиска: точнее указать время появления и редакции документа, язык, доменную область (или даже сайт), формат (например, если ищете xls или ppt). Место на странице (в заголовке или в теле текста), расстояние между поисковыми словами).
- ▶ Булев поиск (логические отношения между поисковыми словами). Сейчас уже чаще не надо писать операторы ("and", "or", "not"), а заполняются поля («без слов»). Хотя при помощи операторов Вы можете соорудить сложный запрос, с несколькими уровнями скобок. Поиск «в найденном».
- ▶ У поисковых машин разные синтаксисы. Смотрите на страницы расширенного поиска и на хелпы:
  - <http://www.yandex.ru/advanced.html>, <http://www.yandex.ru/info/compare.html> (сравнение поисковых машин),
  - <http://www.yandex.ru/info/syntax.html>, <http://www.google.com/intl/ru/help/features.html>, <http://www.yandex.ru/info/syntax.html>, <http://www.google.com/intl/ru/help/features.html>, <http://www.google.ru/intl/ru/help/interpret.html> (экспертиза страницы)

# Для примера: советы по поиску в Яндексе -

[http://www.yandex.ru/ya\\_detail.html](http://www.yandex.ru/ya_detail.html)

- ▶ **Проверяйте орфографию**
- ▶ **Используйте синонимы**
- ▶ **Ищите больше, чем по одному слову**
- ▶ **Не пишите большими буквами**
- ▶ **Найти похожие документы**  
Если один из найденных документов ближе к искомой теме, чем остальные, нажмите на ссылку "найти похожие документы". Ссылка расположена под краткими описаниями найденных документов. Яндекс проанализирует страницу и найдет документы, похожие на тот, что вы указали.
- ▶ **Используйте знаки "+" и "-"**  
"путеводитель по парижу -агентство -тур".
- ▶ **Попробуйте использовать язык запросов**
- ▶ **Искать без морфологии**
- ▶ **Поиск картинок и фотографий**

Была еще полезная вещь – трункация: подстановка вместо неизвестных или любых символов искомого слова (допустим, окончаний) специальных символов (чаще всего – «?» = «один любой символ» и «\*» = «любое количество символов»). Но с реализацией русской морфологии – исчезла?

# 3. Куда вбивать ключевые слова

1. Выделяются три типа ресурсов для поиска в Интернет: поисковая машина, каталог, метапоисковая машина (работающая с поисками других машин).
2. ПМ различаются полнотой проиндексированного Веба и принципами индексации, но главное – алгоритмами ранжирования результатов! Поэтому результаты могут быть очень разными. «Полнота» (min. релевантных документов не попало в результаты поиска) и «точность» (min. нерелевантных попало).
  - Вес, который присваивается заголовкам, ключевым словам, htm-тексту, не-htm-тексту. Полнота индексируемых материалов (заголовки, стартовые страницы и уровни вложенности).
  - Списки «тяжелых», «средних» и «легких» слов и разные тактики отношения к ним при запросе (допустим, Вы ищете о переводах на европейские языки E. Schwarz или «Война и мир»).
  - Выбор разных соотношений между релевантностью страницы и ее популярностью (которая тоже может считаться по-разному).
  - Разная рубрикация (см. ниже, «каталоги»)
3. Еще важнейший критерий выбора ПМ: комфортность ее использования и наличие подходящих для Вас фишек. Но все равно в важном для Вас исследовании лучше пользоваться хотя бы двумя ПМ.

# Примеры фишек Google:

- ▶ Неспециализированные для ученых: подсказки правописания, калькулятор, поиск карт, картинок, товаров, определений, синонимов и т.п. Answers – платные вопросы
- ▶ Специализированные для ученых:
  - <http://books.google.com/> и <http://print.google.com/> - поиск и покупка книг, бумажных и электронных, с возможностью частичного предварительного ознакомления.
  - <http://scholar.google.com/> - поиск только внутри «официального» научного Интернета (журналы и библиотеки с репутацией, кто кого цитирует).
  - Groups - группы новостей (сетевые конференции, конференции, телеконференции). Т.е. тематические обсуждения в Интернете (форумы).
  - News – новости (события), о которых (в т.ч. и давно) писали разные авторы и ресурсы.
- ▶ Высокая степень персонализации: Google desktop (ну очень медленно индексирует!), история поисков, настройка панели виджетов, можно пытаться агрегировать новостные ленты (<http://www.google.com/reader/>) и т.п.
  - [www.google.com/alerts](http://www.google.com/alerts) - можно заказать (ежедневные) оповещения электронной почтой: например, с новостями на определенную тему или, допустим, заставить Гугл регулярно проводить интересующий Вас сложный поиск. Учитывать ли (как и при рассылках) рекламу, замусоривание ящика...

Ситуация на рынке поисковых машин все время меняется!

# 4. Если поиск не принес результата

- ▶ Если хорошие результаты есть, но их недостаточно:
- ▶ Можно поискать кнопки «Еще с сайта», «Показать похожие», «Similar» и т.п.
  - Обязательно стоит посмотреть «карту сайта», который Вам понравился. Контент может оказаться полезным даже при несовпадении ключевых слов.
  - Некоторые из кнопок выше высвечиваются, только если поставить соответствующие настройки поиска.
- ▶ Если результаты вообще не годятся
- ▶ Прежде чем начинать поиск с новыми словами, есть смысл проанализировать результаты экспертизы (<http://www.google.ru/intl/ru/help/interpret.html>, <http://help.yandex.ru/search/?id=481937>). Иногда анализ статистики слов позволяет понять, как можно переформулировать запрос.

- ▶ ПМ могут соответствовать определенные каталоги (<http://list.mail.ru/index.html>) ПМ могут соответствовать определенные каталоги (<http://list.mail.ru/index.html>, <http://yaca.yandex.ru/yaca/cat>) ПМ могут соответствовать определенные каталоги (<http://list.mail.ru/index.html>, <http://yaca.yandex.ru/yaca/cat>).

Не надейтесь на интеллект поисковых машин!

Если, например, поиск принес Вам запись нужного типа, но только одну, вбейте именно ее в поисковую строку – найдите сайты, которые ее упоминают. Думайте!

# Содержательная экспертиза результатов поиска

- ▶ Опубликованному в Интернете опасно доверять:

Ошибочные, базирующиеся на недостаточной, недостоверной или непроверенной информации, устарелые, не учитывающие возможных возражений, с отсутствующей аргументацией, идеологически и коммерчески ангажированные...

- ▶ Вопросы, которые надо поставить сомнительному ресурсу:

Какова цель ведения ресурса, кто отвечает (по содержанию) за ведение сайта (как туда попадает информация), как происходит обновление и правка, **подтверждается ли информация независимыми (т.е. не копирующими информацию друг у друга) ресурсами?**

- ▶ Составить впечатление о ресурсе:

Тон и стилистика, манера обобщений, возможность коммерческой или политической выгоды, манера обращений с позициями оппонентов, наличие и характер цитирования (в т.ч. и онлайн), отношение к копирайту...

Разумеется, надежнее Интернет-представительства оффлайн-ресурсов. Хотя



## 5. Что еще можно найти / искать в Интернет?

► То, что есть в офф-лайне. Полнотекстовые ресурсы

- Электронные библиотеки
- Каталоги офф-лайн библиотек ([www.sigla.ru](http://www.sigla.ru))

В т.ч.:

- <http://liber.rsuh.ru/section.html?id=681> (в т.ч. EBSCO, JSTOR и т.д.)
- <http://www.ub.ruhr-uni-bochum.de/>

► То, что есть только / по преимуществу в офф-лайне

- Ресурсы, связанные с людьми и их деятельностью (применительно к гуманитарным исследованиям – часто бывает очень полезно искать программы курсов, а потом списываться с их участниками).
- Фонды, гранты, конкурсы и т.п.

► То, что есть только в онлайне

- Ресурсы самого Интернет – больше 80 % информации. Рейтинги, ссылки, гостевые, переписка, блоги, чаты, поисковые запросы.
- Рассылки. Критерий (ощущения) принадлежности к научному сообществу - <http://www.h-net.org/lists/> Рассылки. Критерий (ощущения) принадлежности к научному сообществу - <http://www.h-net.org/lists/> (поиск: <http://www.h-net.org/logsearch/>).
- Кэш-копии (сохранившиеся в поисковых системах копии уже снятых страниц)

Хороший сайт: <http://theoryofculture.narod.ru/links/>

# «Видимый» и «невидимый» (неиндексированный) Интернет

Кузнецов С. Как вести бизнес-разведку в «невидимом»  
интернете -  
<http://www.rususa.com/news/news.asp-nid-21056-catid-2>

- ▶ базы данных (в т.ч. библиографические записи) и прочие неформаты
  - ▶ страницы глубокой вложенности
  - ▶ ответы на динамическую информацию (т.е. все, где пользователь что-то должен ввести, простыми поисковиками не индексируется)
  - ▶ с момента прошедшего индексирования (как правило, несколько месяцев)
  - ▶ соотношения между сайтами
- и т.д.

Если чего нет в Интернете, это не означает, что этого не было



# Выбирать – OPAC (Katalogrecherche ohne Anmeldung)

OPAC der Bibliothek der Ruhr-Universität - Microsoft Internet Explorer

Файл Правка Вид Избранное Сервис Справка

Назад Поиск Избранное

Адрес: [http://benutzer.ub.ruhr-uni-bochum.de/scripts/www\\_opac.dll/opacmsk?benutzer=3F693D8C377ACA6A](http://benutzer.ub.ruhr-uni-bochum.de/scripts/www_opac.dll/opacmsk?benutzer=3F693D8C377ACA6A) Переход

**Benutzerarbeitsplatz der Universitätsbibliothek**  
OPAC der Bibliothek der Ruhr-Universität

Bibliotheks-Benutzerarbeitsplatz > Rechenzentrum > zentrale Einrichtungen

**Gemeinsamer OPAC der Bibliotheken der Ruhr-Universität Bochum, der Fachhochschule Bochum, der Evangelischen Fachhochschule Bochum und der Universität Dortmund**

zurück Auswahlmenü

Info zur Suchmaske Start der Suche oder Eingaben löschen Indexanzeige

Suche nach	allen Literaturtypen			
in	allen Bibliotheken			
Stichwort	<input type="text"/>			
Person	<input type="text"/>			
ISBN	<input type="text"/>			
Erscheinungsjahr	von	<input type="text"/>	bis	<input type="text"/>
Trefferanzeige	max	50	ab	1

Info zur Suchmaske Start der Suche oder Eingaben löschen Indexanzeige

Готово Интернет

пуск Входящие - Microsof... 14 Internet Explorer 3 Microsoft Word Microsoft PowerPoint ... EN 22:26

# Ресурс, которым я пользовался при подготовке презентации:

- ▶ Презентация «Conducting Online Research» - <http://owl.english.purdue.edu/owl/resource/558/1> Презентация «Conducting Online Research» - <http://owl.english.purdue.edu/owl/resource/558/1>. Сайт в целом: <http://owl.english.purdue.edu/sitemap/> - много по academic writing, очень полезно также для изучения «функционального английского».
- ▶ Порядка 750 хендаутов и презентаций
  - **The Writing Process**
  - **Professional, Technical, and Job Search Writing**
  - **General Academic Writing**
  - **Research and Citation**
  - **Grammar and Mechanics**
  - **English as a Second Language (ESL)**
  - **Internet Literacy**
  - **Writing in the Social Sciences**
  - **Writing in Engineering**
  - **Literary Analysis and Criticism**
  - **Creative Writing**
  - **Teaching Writing**
  - **Tutoring Writing**
  - **About the OWL at Purdue**

# Вместо резюме: поиск и серфинг

Взаимодополняющие логики нахождения ресурсов: поиск (Вы предписываете машине, что она должна показать) и серфинг (Вы идете по ссылкам, которые предлагает машина).

При серфинге – не забывать о поиске!  
(да и наоборот)

Параллельно поиску (и серфингу) – продолжайте пополнять «Избранное» (на десктопе, в системе закладок, на бумаге). Аннотируйте, обязательно давайте названия.

- Так же, как ведение библиографических записей параллельно чтению.

Как и практически во всех аспектах осуществления (научного) проекта: сосуществование «линейной» и «циклической» временных разверток.

- «Выбираю тему - знакомлюсь с литературой - распиливаю источник - составляю план - пишу – редактирую» vs. «перечитываю свой текст - доподчитываю литературу - переопределяю тему» и т.д. Эта двухфазность должна не повергать в шок, а быть предметом контроля.
- Иными словами: план непременно нарушится, но это нарушение нужно запланировать. Серфинг обязательно отвлечет Вас, но от этого отвлечения нужно будет вернуться.