

# **Информационные базы данных ПО ХИМИИ**

**Александр Антонович  
Рагойша**

Кафедра общей химии и методики  
преподавания химии  
к. 501-а

102 часа, из них аудиторных – **52**

14 недель:

2 лекции (4 час.) +  
+ **12** занятий x **4** час.

(минилекции + практикум + УСП)

Зачет (3 зач. ед.)

# ЛИТЕРАТУРА

- А. А. Рагойша. Поиск химической информации в Интернете. Поисковые системы и тематические каталоги: Учеб. пособие для студентов хим. фак. – Мн.: БГУ, 2003.
- А. А. Рагойша. Поиск химической информации в Интернете: научные публикации : учеб. пособие для студентов хим. фак. спец. 1-31 05 01. – Мн.: БГУ, 2007.
- В. М. Потапов, Э. К. Кочетова. Химическая информация. Где и как искать химику нужные сведения. – М.: Химия, 1988.
- Рагойша, А. А. [Текстовый поиск научной химической информации в Интернете] : практикум по курсу "Информационные технологии в химии" для студентов спец. 1-31 05 01 Химия (по направлениям) — Мн.: БГУ, 2012.  
<http://elib.bsu.by/handle/123456789/14599>
- А. А. Рагойша. Азбука веб-поиска для химиков. – Минск, БГУ, 1999-2020. <http://www.abc.chemistry.bsu.by>

## АЗБУКА ВЕБ-ПОИСКА ДЛЯ ХИМИКОВ Материалы учебных дисциплин "Информационные технологии в химии" и "Информационные базы данных по химии"



### Презентации к лекциям

1 курс  
"Информационные базы данных по химии"  
Лекция 1. Синтаксис запроса



### Практикум

- 1а. Текстовый поиск информации
- 1б. Патентные базы данных
2. Структурный поиск



**ABC-Chemistry** - каталог постоянно доступных полнотекстовых онлайн-журналов и информация о временно доступных журналах



Хімічны часопісы праз кампутарную сетку БДУ. Анлайнвыяны базы дадзеных у бібліятэках Беларусі



25 стартовых точек поиска химической информации



**Суперхімікі** :  
Хімічныя алімпіяды школьнікаў Беларусі : ілюстраваная гісторыя



Напоминаем, что для сотрудников и студентов БГУ через сайт **Фундаментальной библиотеки БГУ** существует свободный доступ к базам данных **Web of Science** (*Science Citation Index* и др.) и **Scopus**.

В числе иных полезных ресурсов, доступных через библиотеку БГУ, упомянем:

- Несколько российских журналов на платформе **eLIBRARY.RU**
- Информационный комплекс **EBSCO**, в том числе, базу данных **Academic Search Premier**.
- Евразийскую патентную информационную систему **ЕАПАТИС** и др.

Временно доступны через сеть БГУ:

- **SpringerLink** - Springer eBooks (2016-2021) и Springer journals (1997-2022).

Напоминаем об иных бесплатных ресурсах:

**Тематические спецвыпуски** журналов издательства *Springer*.  
**ChemRxiv**: The Preprint Server for Chemistry - архив препринтов по химической тематике (совладельцы: *ACS, RSC, German Chemical Society GDCh, Chinese Chemical Society CCS, Chemical Society of Japan CSJ*).  
**ChemRN** - Chemistry Research Network - новый раздел архива препринтов SSRN (с 2016 г. собственность издательства *Elsevier*).

ENHANCED BY Google

Search



# Указатели ресурсов

## Поисковая система - Search engine

- робот (паук)
- индекс (база данных)
- поисковая программа, веб-интерфейс

Универсальные: Google, Яндекс, Bing, ...

Специализированные (**вертикальный поиск**)

**Тематический каталог (Directory)**

- **Метапоисковая система**  
использует индексы нескольких иных поисковых систем
- **Специализированная база данных**  
(робот отсутствует)
- **Метасайт** -  
небольшой по объему сборник ссылок на веб-страницы

**Browse (перелистывание) -- Search (поиск)**

**Браузер            Mozilla Firefox, Google Chrome**

# Синтаксис запроса в текстовых базах данных

- База данных (database) -  
упорядоченный информационный массив,  
состоящий из стандартных блоков.

Классификация по типу содержимого:

текстовые,  
числовые,  
формульные,  
...

# Структура базы данных (с точки зрения пользователя)

- **Запись (record)** - стандартный блок информации

- **Поле (field)** - смысловой фрагмент записи

Поля:  
текстовые,  
числовые  
и др.

Поле	Значение
Заглавие	Химия сегодня и завтра
Издано в	Мн. : Университетское, 1987.
Примечания	Библиогр.: с. 126-127 (42 назв.). 9630 экз.
Тематика	ХИМИЯ ХІМІЯ
УДК	54
ГРНТИ	31.01
Автор (лицо/организация)	Свиридов, Вадим Васильевич

*Запись в каталоге библиотеки*

- Поисковая программа  
(search and retrieval software)



имеет страницу с **поисковым бланком**,  
предназначенным для формулирования  
запроса

- **Запрос (query)** -  
поисковое задание, содержащее поисковые термины  
и инструкцию по их интерпретации программой

Пример запроса:

**натрий**

Заполняем  
поисковый бланк:

Поисковая программа ищет в своей базе данных те записи, в которых присутствует слово **натрий**

Список  
обнаруженных  
записей  
выводится на  
экран

[Натрий](#) - [ [Translate this page](#) ]

**Натрий** - жизненноважный межклеточный и внутриклеточный элемент, участвующий в ...  
Потребность в **натрии** минимально составляет около 1 г/сут и в значительной ...  
[www.sunduk.ru/Encycl/ChemFood/C027.htm](http://www.sunduk.ru/Encycl/ChemFood/C027.htm) - [Cached](#) - [Similar](#)

[НАТРИЙ](#) - [ [Translate this page](#) ]

**Натрий-22** с периодом полураспада 2,58 года используют в качестве источника позитронов. **Натрий-24** (его период полураспада около 15 часов) применяют в ...  
[www.krugosvet.ru/enc/nauka\\_i\\_tehnika/.../NATRI.html](http://www.krugosvet.ru/enc/nauka_i_tehnika/.../NATRI.html) - [Cached](#) - [Similar](#)

**Поиск - не по смыслу, а по факту наличия термина!**

Нет стандартного синтаксиса запроса.

У каждой программы **свои** правила.

**Иногда** правила совпадают  
(но необязательно, что полностью).

Бывает, что некоторые элементы  
разными поисковыми программами  
воспринимаются *с точностью до наоборот*.

# Логические (Булевы) операторы

- **AND**  
натрий AND калий

& , ...

- **OR**  
натрий OR калий

| , ...

- **NOT**  
натрий NOT калий

- , (andnot, and not, but not)

варианты  
обозначений

## Оператор по умолчанию (default operator)

*Пример:* Обе записи равнозначны, если AND – по умолчанию:

натрий AND калий

натрий калий

# Порядок выполнения операций

- Сначала: NOT и AND, затем: OR
- Если нужно, порядок меняют круглыми скобками

*Пример:*

Найти записи, в которых:  
обязательно присутствует **натрий** или **калий** и  
обязательно присутствует **фосфат** или **силикат**

**Правильно:**

(натрий OR калий) AND (фосфат OR силикат)

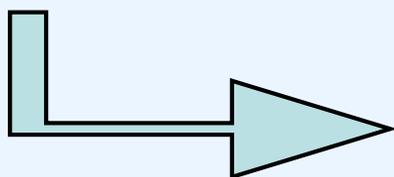
**Неправильно:**

натрий OR **калий AND фосфат** OR силикат

# Операторы расстояния - 1

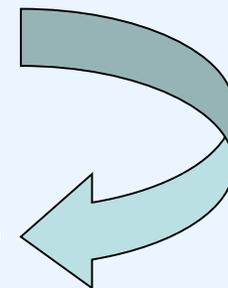
- Кавычки

Пример: "фосфат натрия"



два алгоритма:  
**фраза** из 2 слов *или*  
**строка** из 13 символов

"фосфат\_натрия"  $\neq$  "фосфат\_\_натрия"

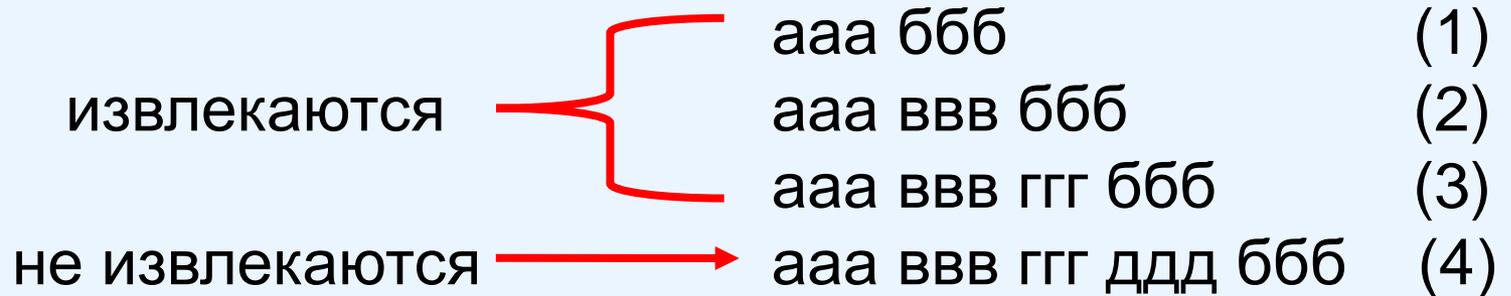


*(символом подчеркивания обозначен пробел)*

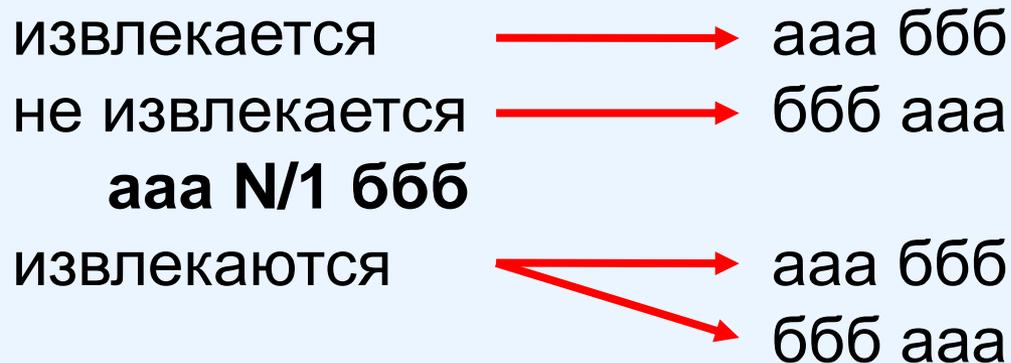
# Операторы расстояния - 2

- WITH/n , NEAR/n (W/n, N/n, WITH, ...)
- n – расстояние *или* число посторонних слов

Пример: **aaa WITH/3 ббб**



Пример: **aaa W/1 ббб**



# Шаблон - 1

\* ("звездочка")

заменяет **любое число** символов ( в т. ч. нулевое)

*Примеры:* **хлор\***

хлор, хлорид, хлораты, хлорирование, ...

**НО:** хлорофилл

**\*фосфат**

фосфат, **ди**фосфат, **поли**фосфат, ...

Wildcard. Truncation (right-hand, left-hand) - Усечение

# Шаблон - 2

? (вопросительный знак), # (решетка)  
заменяет **ОДИН** СИМВОЛ

*Пример:* **бут?н**  
бут**а**н, бут**е**н, бут**и**н, бут**о**н

*Как правило:*

При шаблоне оставлять не менее трех букв.  
Не использовать шаблон внутри кавычек.

Шаблон увеличивает количество  
информационного мусора в результатах поиска

# Stemming

- **Stemming** – режим работы поисковой программы, при котором происходит **учет грамматических форм** терминов (**учет морфологии, учет словоформ**)

*Пример:* **фосфат** 

фосфат, фосфатами, фосфатный, ... (полифосфат - ?)

*Пример:* **write** 

write, writes, writing, wrote

**Не** проводить stemming:

**"фосфатами"**

# Стоп- слова

- **Стоп-слова (stopwords)** - слова, которые при поиске не учитываются.

Это слова, не несущие самостоятельной смысловой нагрузки, но особенно часто встречающиеся в тексте:  
предлоги, союзы, артикли и т. п.

*Пример:*

~~**The Analyst**~~

Включить стоп-слово в поиск:

**"The Analyst"**

# Регистр букв

- Абсолютное большинство поисковых программ нечувствительно к регистру букв – для них **строчные и заглавные** буквы в запросе **равнозначны**.

*Пример:*

**фосфат AND силикат**

**фосфат and силикат**

**фОсФаТ aNd СиЛиКаТ**

*годится любой вариант*

## Указание поля поиска

- Поиск можно сделать более эффективным, если проводить его не по записям в целом, а только по избранным полям.

Для этого в запросе рядом с поисковым термином указывают код соответствующего поля.

Коды полей в разных базах данных – разные.

*Примеры:*

**ttr/фосфат**

**фосфат filetype:pdf**

# Особенности ПОИСКОВЫХ СИСТЕМ

# Логические операторы

- Операторы отображать **ЗАГЛАВНЫМИ** буквами  
(если строчными – стоп-слово)

- **AND, +, &**

натрий AND калий  
натрий & калий  
натрий +калий

Обычно достаточно:  
**натрий калий**

- **OR, |, ...**

натрий **OR** калий  
натрий | калий

- **NOT, -**

натрий NOT калий  
натрий **-калий**

без  
пробела!

# Регистр букв

- Нечувствительны к регистру -  
большинство

## Шаблон

Google, Яндекс

Звездочка \* заменяет собой любое **слово**.  
Используют внутри кавычек, *например*:

**"особенно полезен \* чай"**

## Поиск по словоформам

Google – в основном, нет.

Яндекс – да.

## Поиск по полям документов

Возможен, но полезность ограниченная  
основная часть информации на веб-странице  
находится в одном самом большом поле

# Google: поиск по полям - 1

- **intitle:БГУ**
- **intitle:БГУ химфак**
- **intitle:БГУ intitle:химфак**  
**allintitle:БГУ химфак**

intitle:БГУ химфак

About 6,990 results (0.31 seconds)

[Химический факультет БГУ - Главная страница](#)  - [www.chemistry.bsu.by/](#) - *Cached*

Информация о факультете: деканат, кафедры, лаборатории, ко программы.

<a href="#">Абитуриенту</a>	<a href="#">Кафедра не</a>
<a href="#">www.chemistry.bsu.by/...</a>	<a href="#">www.chemis</a>
<a href="#">/index.asp?...</a>	<a href="#">/index.asp?</a>
В день этого праздника <b>химфак</b>	электронна

allintitle:БГУ химфак

About 989 results (0.06 seconds)

[БГУ ХИМФАК | Химия и Химики | БГУ ХИМФАК](#)  - [ [Translate this page](#) ]

[www.himiki.com/bgu-ximfak](#) - *Cached*

Белорусский государственный университет Химический факультет Факультет.

# Google: поиск по полям - 2

- **site:bsu.by**
- **site:bsu.by кислород**

[Операция "Кислород" - Энциклопедия кораблей](#) 🔍 - [ Translate this page ]  
[ship.bsu.by/text.aspx?guid=4024](http://ship.bsu.by/text.aspx?guid=4024) - Cached

Операция "Кислород" проведена французскими боевыми пловцами 10 июля 1985 года, в результате которой в новозеландском порту Окленд судно ...

[Открытая студенческая олимпиада 'Кислород-2010'](#) 🔍 - [ Translate this page ]  
[www.abc.chemistry.bsu.by/oxygen/default.htm](http://www.abc.chemistry.bsu.by/oxygen/default.htm) - Cached

19-20 июня 2010 г. на химическом факультете БГУ прошла Первая открытая студенческая олимпиада по химии "Кислород-2010". В состязаниях ...

# Google: поиск по полям - 3

- **filetype:pdf** **ext:pdf**

filetype:pdf intitle:БГУ химический факультет

filetype:pdf intitle:БГУ химический факультет|

6 results (0.26 seconds)

[PDF] [ПРОГРАММА ФЕСТИВАЛЯ ФАКУЛЬТЕТОВ БГУ](#) Место проведения: В

- [ [Translate this page](#) ]

[www.ffsn.bsu.by/ffsn.files/news/prog\\_fest\\_2011.pdf](http://www.ffsn.bsu.by/ffsn.files/news/prog_fest_2011.pdf)

File Format: PDF/Adobe Acrobat - [Quick View](#)

16.00-16.30 **Химический факультет**. 16.30-17.00 Биологический **факультет**. 17.00-17.30

**Факультет** доуниверситетского образования. 17.30-18.00 Военный **факультет** ...

# Google: поиск по полям - 4

- **inurl:bsu**

inurl:bsu inurl:chemistry

[BSU, Faculty of Chemistry](#) 🔍

[wikimapia.org/3360176/BSU-Faculty-of-Chemistry](https://wikimapia.org/3360176/BSU-Faculty-of-Chemistry) - Cached  
World / Belarus / Minsk / Minsk , 2km from center (Минск  
27°32'57"E. BSU, Faculty of Chemistry (Minsk) ...

[www.Chemistry.bsu.by](http://www.Chemistry.bsu.by) 🔍

[www.markosweb.com/www/chemistry.bsu.by/](http://www.markosweb.com/www/chemistry.bsu.by/) - Cached

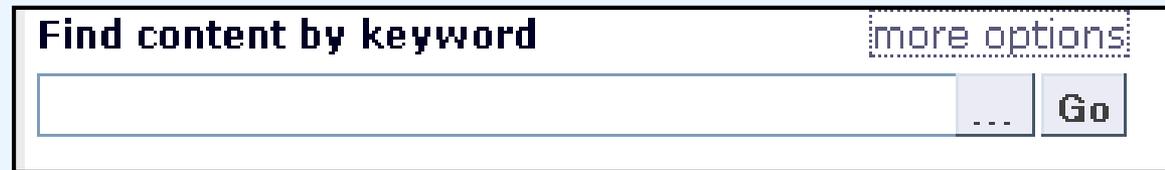
# Поисковый бланк и список результатов поиска

# Поисковый бланк - 1

- Простейший, Basic, Quick, Simple
- Расширенный, Advanced, Expert

Классификация условна – в рамках определенной базы данных.

Обычно такие бланки называют **Quick Search**



The image shows a search form with a title bar that reads "Find content by keyword". To the right of the title bar is a link labeled "more options". Below the title bar is a large text input field. To the right of the input field are two buttons: a small grey button with three dots "..." and a larger grey button labeled "Go".

## Поисковый бланк - 2

*Пример бланка*

Quick Search:

Поиск по двум разным полям; использование булевых операторов; выбор временного интервала.

Query <a href="#">[Help]</a>			
Term 1:	<input type="text"/>	in Field 1:	<input type="text" value="Title"/>
			<input type="button" value="v"/>
		<input type="text" value="AND"/>	<input type="button" value="v"/>
Term 2:	<input type="text"/>	in Field 2:	<input type="text" value="Inventor Name"/>
			<input type="button" value="v"/>
Select years <a href="#">[Help]</a>			
	<input type="text" value="1976 to present [full-text]"/>		<input type="button" value="v"/>
		<input type="button" value="Search"/>	<input type="button" value="Reset"/>

# Поисковый бланк - 3

Пример бланка

Advanced Search:

All Sources	Journals	Books	Advanced Search
Term(s):	<input type="text"/>	within:	Title <input type="button" value="v"/>
<input type="button" value="AND"/> <input type="button" value="v"/>	<input type="text"/>	within:	Authors <input type="button" value="v"/>
Include:	<input checked="" type="checkbox"/> Journals	<input checked="" type="checkbox"/> All Books	
Source:	<input type="text" value="All sources"/> <input type="button" value="v"/>		
Subject:	Select one or more:		
	<input type="button" value="v"/> - All Sciences - <input type="button" value="^"/>		Hold down the key) to select m
	Agricultural and Biological Sciences <input type="button" value="≡"/>		
	Arts and Humanities		
	Biochemistry, Genetics and Molecular Biology <input type="button" value="v"/>		
Dates:	<input checked="" type="radio"/> All Years	<input type="radio"/> 1999 <input type="button" value="v"/>	to: Present <input type="button" value="v"/>
<input type="button" value="Search"/> <input type="button" value="Clear"/> <input type="button" value="Recall Search"/>			

# Элементы бланка

The image shows a search form with the following elements:

- Navigation tabs: **All Sources** (highlighted), **Journals**, **Books**
- Term(s):
- AND
- Include:  Journals  All Books
- Source:
- Select one or more:
- Subject: 
  - Agricultural and Biological Sci
  - Arts and Humanities
  - Biochemistry, Genetics and M
- Dates:  All Years  1999  to:
- Buttons: **Search**, **Clear**, **Recall Sea**

- Графа бланка (редактируемая графа, редактируемое поле).
- Список.
- Меню (выпадающий список).
- Переключатель.
- Выключатель.
  
- Текстовые пояснения.
- Ссылка на иной бланк.
- Ссылка на блок инструкций.
- Кнопка начала поиска.

## Advanced Search

Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from:

to

# Список результатов поиска

Пользователь получает результаты поиска в форме списка обнаруженных документов.

Список может быть сформирован:

- по алфавиту,
- в хронологическом порядке  
(прямом или обратном),
- по релевантности.

# Релевантность

Релевантность документа –  
степень соответствия его поисковому заданию.

При расчете релевантности учитываются:

- количество поисковых терминов в документе,
- расстояние между ними в тексте,
- их плотность,
- их порядок расположения,
- их место – в названии, в начале записи или в конце,
- и др.

# Особенности ПОИСКОВЫХ СИСТЕМ

# Особенности работы

Предназначены для широкой публики

Упростить взаимодействие с системой

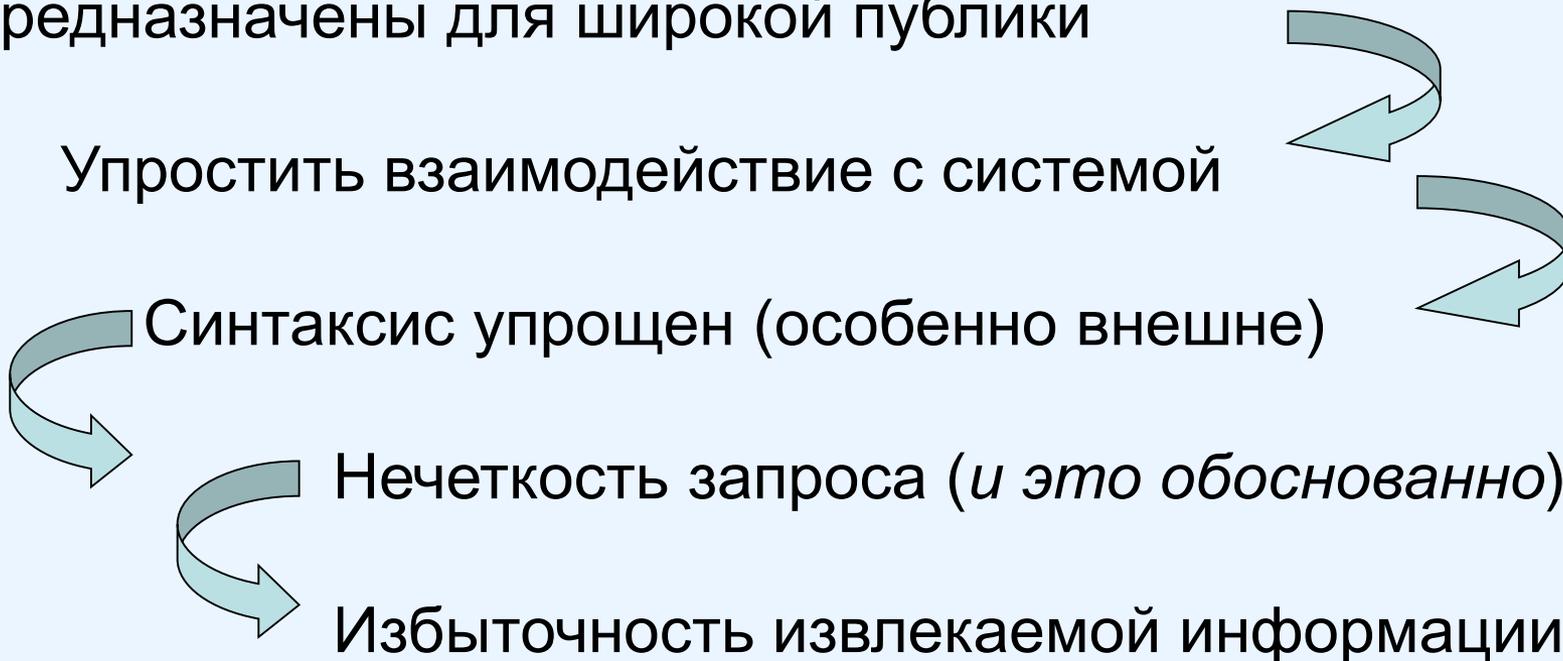
Синтаксис упрощен (особенно внешне)

Нечеткость запроса (*и это обоснованно*)

Избыточность извлекаемой информации

Огромный список результатов поиска

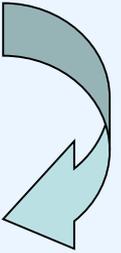
Главное – не величина списка, а **качество сортировки**



# Релевантность - 2

Расчет **ранга** веб-страницы – т.е. ее авторитетности

Сколько ссылок с других страниц на данную?  
Каков ранг ссылающихся страниц?



Расчет **релевантности**:

+ Вклад 1: поисковые термины на странице  
+ Вклад 2: ранг страницы

---

= Релевантность страницы в списке результатов поиска

# О достоверности информации

## Традиционная vs. онлайн

- **Печатная литература**  
автор известен  
контроль со стороны издателя
- **Научная литература**  
система **рецензирования** (peer review)
- **Веб-источники**  
анонимность, отсутствие контроля – почти норма

**Достоверность информации лежит в широких пределах:**  
от объективной - до субъективной,  
от полностью достоверной - до ложной  
и до намеренно сфальсифицированной

# Оценка ресурса

В основе оценки онлайн-источника лежат известные критерии оценки печатных источников:

Репутация автора;

Наличие контроля качества;

Степень объективности изложения;

Признаки актуальности *или* даты создания.

Является ли первоисточником?

## Плюс веб-специфика:

- *Рекламные* блоки могут казаться частью документа.
- *Отсканированный* и оптически распознанный текстовый материал редко выверяется корректорами.
- Содержание веб-страницы может быть изменено *несанкционированно* (атака хакера, прихоть администратора).
- Проблемы субъективности/достоверности особенно остро проявляются в *форумах, блогах, социальных сетях*.

# Стиль

## Лингвистика

Явные признаки низкокачественного ресурса:

- Обилие опечаток и грамматических ошибок.
- Развязный стиль изложения.

## Дизайн

Эксперт тщательно оценивает содержание, а обычный потребитель больше доверяет внешнему виду страницы.

# Формальный анализ URL

## Доменное имя

достоверность выше:

.gov .edu .ac.uk . ac.jp

достоверность ниже:

livejournal.com

## Папки

повысить бдительность:

~... private, members

# Предпочтительны

## Сайты:

- университетов,
- научных обществ,
- научных издательств,
- официальных патентных бюро,
- авторитетных коммерческих организаций,
- персональные сайты ученых.

Стремимся работать с **первоисточниками**  
и интенсивно используем **свой** мозг