

Направление «Филология»

Профиль:

«Компьютерная лингвистика»

КОД - 270

Время выполнения задания 180 минут

Задание 1.

Решите задачу:

В столбец написано 4 слова. Всегда ли можно вычеркнуть одно так, чтобы оставшиеся 3 шли или в прямом или в обратном алфавитном порядке? Ответьте на аналогичный вопрос, если написано 5 слов, а вычеркнуть надо 2.

Задание 2.

Перед вами стоит задача разработать систему для дигитализации 86 томов Энциклопедического Словаря Ф.А.Брокгауза и И.А.Ефрона. Вам нужно не просто оцифровать Словарь, но представить его содержимое максимально полно в виде открытой базы данных, интегрированной с википедией. Примеры страниц представлены в приложении.

Ваша задача состоит в том, чтобы:

- а) Составить мини-проектное задание, в котором были бы сформулированы основные характеристики конечного продукта: те, которые вам кажется необходимыми, а также те, которые вы считаете желательными. Мотивируйте ваше решение: какие возможности предоставляют пользователю те или иные характеристики продукта?
- б) Определить основные этапы проекта и их последовательность.
- в) Составить список лингвистических проблем, с которыми вы, возможно, столкнетесь на каждом из этапов, и предложить пути их решения.

Задание 3.

Прочтите пост из блога, посвященного автоматической обработке естественного языка (<http://nlpers.blogspot.com/>). Сформулируйте основные идеи, связанные с задачами компьютерной лингвистики, высказанные автором в данном посте. Напишите краткое эссе объемом не более двух страниц на русском языке о том, как вы понимаете эти идеи. Если у вас возникли собственные соображения по этой теме, выскажите их.

The friend I crashed with while attending SODA is someone I've known since we were five years old. (Incidentally, there's actually someone in the NLP world who I've actually known from *earlier*...small world.) Anyway, the friend I stayed with is just finishing med school at

UCSF and will soon be staying there for residency. His specialty is neurosurgery, and his interests are in neural pathologies. He spent some time doing research on Alzheimer's disease, effectively by studying mice (there's something I feel sort of bad about finding slightly amusing about mice with Alzheimer's disease). Needless to say, in the process of doing research, he made nearly daily use out of PubMed. (For those of you who don't know, PubMed is like the ACL anthology, but with hundreds of thousands of papers, with new ones being added by the truckload daily, and will a bunch of additional things, like ontologies and data sets.)

There are two things I want to talk about regarding PubMed. I think both of these admit very interesting problems that we, as NLPers, are qualified to tackle. I think the most important thing, however, is opening and maintaining a *wide* channel of communication. There seems to be less interaction between people who do (for instance) bio-medical informatics (we have a fairly large group here) and what I'll term as mainstream NLPers. Sure, there have been BioNLP workshops at ACLs, but I really think that both communities would be well-served to interact more. And for those of you who don't want to work on BioNLP because it's "just a small domain problem", let me assure you: it is *not* easy... don't think of it in the same vein as a true "sublanguage" -- it is quite broad.

I suppose I should give a caveat that my comments below are based on a sample size of one (my friend), so it may not be totally representative. But I think it generalizes.

Search in PubMed, from what I've heard, is good in the same ways that web search is good and bad in the same ways that web search is bad. It is good when you know what you're looking for (i.e., you know the name for it) and bad otherwise. One of the most common sorts of queries that my friend wants to do is something like "show me all the research on proteins that interact in some way with XXX in the context of YYY" where XXX is (eg) a gene and YYY is (eg) a disease. The key is that we don't know which proteins these are and so it's hard to query for them directly. I know that this is something that the folks at Penn (and probably elsewhere) are working on, and I get the impression that a good solution to this problem would make lots and lots of biologists much happier (and more productive). One thing that was particularly interesting, however, is that he was pretty averse to using *structured* queries like the one I gave above. He effectively wants to search for "XXX YYY" and have it realize that XXX is a gene, YYY is a disease, and that it's "obvious" that what he wants is proteins that interact with (or even, for instance, pathways that contain) XXX in the context of disease YYY. On the other hand, if YYY were another gene, then probably he's be looking for diseases or pathways that are regulated by both XXX and YYY. It's a bit complex, but I don't think this is something particularly beyond our means.

The other thing I want to talk about is summarization. PubMed actually archives a fairly substantial collection of human-written summaries. These fall into one of two categories. The first, called "systematic reviews" are more or less what we would think of as summaries. However, they are themselves quite long and complex. They're really not anything like sentence extracts. The second, called "meta analyses" are really not like summaries at all. In a meta

analysis, an author will consider a handful of previously published papers on, say, the effects of smoking on lifespan. He will take the data and results published in these individual papers, and actually do novel statistical analyses on them to see how well the conclusions hold.

From a computational perspective, the automatic creation of meta analyses would essentially be impossible, until we have machines that can actually run experiments in the lab. "Systematic reviews", on the other hand, while totally outside the scope of our current technology, *are* things we could hope to do. And they give us lots of training data. There are somewhere around ten to forty *thousand* systematic reviews on PubMed, each about 20 pages long, and each with references back to papers, almost all of which are themselves in PubMed. Finding systematic reviews older than a few years ago (when they began being tagged explicitly) has actually sprouted a tiny cottage industry. And PubMed nicely makes all of their data available for download, without having to crawl, something that makes life much easier for us.

My friend warns that it might not be a good idea to use *all* systematic reviews, but only those from top journals. (They tend to be less biased, and better written.) However, in so far as I don't think we'd even have hope of getting something as good as a systematic review from the worst journal in the world, I'm not sure this matters much. Maybe all it says is that for *evaluation*, we should be careful to evaluate against the top.

Now, I should point out that people in biomedical informatics have actually been working on the summarization problem too. From what I can tell, the majority of effort there is on rule-based systems that build on top of more rule-based systems that extract genes, pathways and relations. People at the National Library of Medicine, Rindfleisch and Fiszman, use SemRep to do summarization, and they have tried applying it to some medical texts. Two other people that I know are doing this kind of work are Kathy McKeown and Greg Whalen, both at Columbia. The Columbia group has access to a medically informed NLP concept extractor called MedLEE, which gives them a leg up on the low-level processing details. If you search for 'summarization medical OR biomedical' in GoogleScholar, you'll get a raft of hits (~9000).

Now, don't get me wrong -- I'm not saying that this is easy -- but for summarization folks who are constantly looking for "natural artifacts" of their craft, this is an enormous repository.

[\[Библиотека «Вехи»\]](#)

Энциклопедический Словарь Ф.А.Брокгауза и И.А.Ефрона

(В 86 томах с иллюстрациями и дополнительными материалами)

[\[Титульная страница\]](#) | [\[А\]](#) | [\[Б\]](#) | [\[В\]](#) | [\[Г\]](#) | [\[Д\]](#) | [\[Е\]](#) | [\[Ж\]](#) | [\[З\]](#) | [\[И\]](#) | [\[Й\]](#) | [\[К\]](#) | [\[Л\]](#) | [\[М\]](#) | [\[Н\]](#) | [\[О\]](#) | [\[П\]](#) | [\[Р\]](#) | [\[С\]](#) | [\[Т\]](#) | [\[У\]](#) | [\[Ф\]](#) | [\[Х\]](#) | [\[Ц\]](#) | [\[Ч\]](#) | [\[Ш\]](#) | [\[Щ\]](#) | [\[Ы\]](#) | [\[Э\]](#) | [\[Ю\]](#) | [\[Я\]](#)

- [Бехер Зигфрид](#)
- [Бехер Иоганн-Иоаким](#)
- [Бехин](#)
- [Бехлик](#)
- [Бехман](#)
- [Бехоянис-хеви](#)
- [Бехтеевка](#)
- [Бехтеевы](#)
- [Бехтемир-ишан](#)
- [Бехтерев](#)
- [Бехтольсгейм](#)
- [Бехштейн Иоанн-Матвей](#)
- [Бехштейн Людвиг](#)
- [Бехштейн Рейнгольд](#)
- [Бехштейн Рейнгольд \(дополнение к статье\)](#)
- [Бехштейн Фридрих-Вильг.-Карл](#)
- [Бехштейн Фридрих-Вильгельм-Карл \(дополнение к статье\)](#)
- [Бехынька](#)
- [Бец](#)
- [Бецберг](#)
- [Бецирк](#)
- [Бециг](#)

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)

Бехштейн Рейнгольд

(Bechstein) — германист, сын Людвига Б., родился 12 октября 1833 г. в Мейнингене, изучал филологию, но преимущественно немецкий яз. и германские древности. После † отца Б. занял его место библиотекаря в Мейнингене, затем переселился в Иену, был профессором при тамошнем университете, с 1871 г. занимает кафедру немецкой и новейшей литературы в Ростоке. Подобно отцу, он издал разные произведения древнегерманской литературы: "Heinrich u. Kunig unde v. Eberhard v. Erfurt" (1860); продолжал издававшийся отцом "Deutsches Museum" (Neue Folge, 1 т., Лейпц., 1862); составил сборник "Altdeutsche M ä rchen, Sagen und Legenden" (Лейпц., 1863) и издал "Tristan und Isolde in deutschen Dichtungen der Neuzeit" (1876) и др.

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)

mail.ru РЕЙТИНГ
8315854 2203 1391

Я 1200

2392902

[\[Библиотека «Вехи»\]](#)

Энциклопедический Словарь Ф.А.Брокгауза и И.А.Ефрона

(В 86 томах с иллюстрациями и дополнительными материалами)

[\[Титульная страница\]](#) | [\[А\]](#) | [\[Б\]](#) | [\[В\]](#) | [\[Г\]](#) | [\[Д\]](#) | [\[Е\]](#) | [\[Ж\]](#) | [\[З\]](#) | [\[И\]](#) | [\[Й\]](#) | [\[К\]](#) | [\[Л\]](#) | [\[М\]](#) | [\[Н\]](#) | [\[О\]](#) | [\[П\]](#) | [\[Р\]](#) | [\[С\]](#) | [\[Т\]](#) | [\[У\]](#) | [\[Ф\]](#) | [\[Х\]](#) | [\[Ц\]](#) | [\[Ч\]](#) | [\[Ш\]](#) | [\[Щ\]](#) | [\[Ъ\]](#) | [\[Э\]](#) | [\[Ю\]](#) | [\[Я\]](#)

- [Бехер Зигфрид](#)
- [Бехер Иоганн-Иоаким](#)
- [Бехин](#)
- [Бехлик](#)
- [Бехман](#)
- [Бехоянис-хеви](#)
- [Бехтеевка](#)
- [Бехтеевы](#)
- [Бехтемир-ишан](#)
- [Бехтерев](#)
- [Бехтольсгейм](#)
- [Бехштейн Иоанн-Матвей](#)
- [Бехштейн Людвиг](#)
- [Бехштейн Рейнгольд](#)
- [Бехштейн Рейнгольд \(дополнение к статье\)](#)
- [Бехштейн Фридрих-Вильг.-Карл](#)
- [Бехштейн Фридрих-Вильгельм-Карл \(дополнение к статье\)](#)
- [Бехынька](#)
- [Бец](#)
- [Бецберг](#)
- [Бецирк](#)

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)

Бехштейн Рейнгольд (дополнение к статье)

(Bechstein) — германист; ум. в 1894 г.

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)



Ожидание counter.yadro.ru...



Википедия
Свободная энциклопедия

- Заглавная страница
- Рубрикация
- Указатель А—Я
- Избранные статьи
- Случайная статья
- Текущие события

- Участие
 - Сообщить об ошибке
 - Портал сообщества
 - Форум
 - Свежие правки
 - Новые страницы
 - Справка
 - Пожертвования

Печать/экспорт

Инструменты

- В других проектах
 - Викисклад

- На других языках
 - العربية
 - Беларуская
 - Deutsch
 - English

Статья Обсуждение Читать Править Править исходный текст История Поиск

Бехштейн, Людвиг

Материал из Википедии — свободной энциклопедии

[Связать?](#)
[\[править\]](#) [\[править исходный текст\]](#)

Людвиг Бехштейн (нем. *Ludwig Bechstein*; 24 ноября 1801, Веймар, — 14 мая 1860, Майнинген) — немецкий писатель, известный составитель сборников народных сказок.

Содержание [убрать]
1 Жизнь и творчество
2 Наиболее важные труды
3 Литература
4 Ссылки

Жизнь и творчество [править | править исходный текст]

Л. Бехштейн был внебрачным сыном Иоганны Каролины Доротеи Бехштейн и французского эмигранта Луи Дюпонтро. В 1810 году его усыновил дядя, Иоганн Маттеус Бехштейн. После окончания гимназии в 1821 году юноша поступает в ученики к аптекарю в Арнштадте (до 1824 года), затем до 1826 года он — помощник аптекаря в Майнингене и до 1828 года — провизор в Бад-Зальцунгене. Получив стипендию от герцога Саксен-Майнингена Бернгарда II, Л. Бехштейн изучает философию, историю и литературу в Лейпциге (1829) и в Мюнхене (1830). В Лейпциге он вступает в студенческое братство. В 1831 году молодой человек назначается герцогским библиотекарем в Майнингене и в 1833 году — директором герцогской библиотеки. В 1832 году Л. Бехштейн основывает «Хеннебергское общество древней истории», председателем которого остаётся до 1857 года. В 1834 году он пишет и издаёт «Хронику города Майнинген 1676—1834». В 1840 году Л. Бехштейну присваивается чин придворного советника. В 1842 году он вступает в *масонскую ложу* «Шарлотта у трёх гвоздик». В 1848 году он становится руководителем общественным хеннебергским архивом (графства Хеннеберг).

Л. Бехштейн писал как поэтические произведения (лирического и патриотического характера), так и исторические новеллы и романы (роман «Тёмный граф» (Der Dunkelgraf, 1854), однако подлинную славу принесли ему изданные им сборники немецких народных сказок. Ещё в 1823 году он выпускает небольшой сборник «Тюрингские народные сказки». В 1845 году выходит в свет составленная им «Немецкая книга сказок». Вышедшая в 1853 году его «Книга немецких саг» была не столь популярна, как предыдущие сборники сказок, однако в научном мире считается классическим собранием германского эпического творчества. Л. Бехштейн был также автором нескольких путеводителей и изданий для путешественников по Тюрингии.

Людвиг Бехштейн
Ludwig Bechstein



Дата рождения:	24 ноября 1801
Место рождения:	Веймар
Дата смерти:	14 мая 1860 (58 лет)
Место смерти:	Майнинген
Гражданство:	 Германия

[\[Библиотека «Вехи»\]](#)

Энциклопедический Словарь Ф.А.Брокгауза и И.А.Ефрона

(В 86 томах с иллюстрациями и дополнительными материалами)

[\[Титульная страница\]](#) | [\[А\]](#) | [\[Б\]](#) | [\[В\]](#) | [\[Г\]](#) | [\[Д\]](#) | [\[Е\]](#) | [\[Ж\]](#) | [\[З\]](#) | [\[И\]](#) | [\[Й\]](#) | [\[К\]](#) | [\[Л\]](#) | [\[М\]](#) | [\[Н\]](#) | [\[О\]](#) | [\[П\]](#) | [\[Р\]](#) | [\[С\]](#) | [\[Т\]](#) | [\[У\]](#) | [\[Ф\]](#) | [\[Х\]](#) | [\[Ц\]](#) | [\[Ш\]](#) | [\[Щ\]](#) | [\[Ы\]](#) | [\[Э\]](#) | [\[Ю\]](#) | [\[Я\]](#)

[Библиомания](#)

[Библиопол](#)

[Библиос или папирус](#)

[Библиофа](#)

[Библиотека](#)

[Библиотека для чтения](#)

[Библиотека русских и иностранных](#)

[писателей](#)

[Библиотека название журналов](#)

[Библиотекарь](#)

[Библиотековедение](#)

[Библиотекография](#)

[Библиотекономия](#)

[Библиофил](#)

[Библист](#)

[Библия](#)

[Библия бедных](#)

[Библия Рафаэля](#)

[Библия у славян](#)

[Библос](#)

[Бибра](#)

[Бибра](#)

[Бибракта](#)

[Бибрих](#)

[Библиографический словарь](#)

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)

Библиотекарь

— звание, которое встречается уже в древнейшие времена; мы видим библиотекарей в Египте, при дворе Птолемеев; в настоящее время Б. называется обыкновенно лицо, которому поручается заведование книгохранилищем или какой-либо его частью; в некоторых странах, однако, как, напр., в Германии, Б. называется исключительно то лицо, которому поручено управление всей библиотекой; лица же, заведующие отделениями библиотеки, носят разные другие названия: в Германии — Custos, Assistent, во Франции — Conservateur и т. д. С развитием библиотечного дела мало-помалу определилась та сумма качеств и знаний, которая необходима библиотекарю для успешного ведения дела; возникла наука библиотековедения, изучить которую в настоящее время можно только практически — службой в библиотеке же. От библиотекаря требуется способность к усидчивому труду, любовь к порядку, он должен знать новые языки и быть энциклопедически образованным человеком; необходимо также знание всего, что касается истории книги, книгопечатания и т. п. См. Gr äsel, "Grundzüge der Bibliotheksl ehre" (Лейпциг, 1890).

[\[Энциклопедический Словарь\]](#) | [\[Библиотека «Вехи»\]](#)

mail.ru РЕНТИНГ 8315872 2221 1398 Я 1200 2392907

Википедия

Свободная энциклопедия

Заглавная страница
Рубрикация
Указатель А—Я
Избранные статьи
Случайная статья
Текущие события

Участие
Сообщить об ошибке
Портал сообщества
Форум
Свежие правки
Новые страницы
Справка
Пожертвования

Печать/экспорт
Инструменты

На других языках
العربية
Беларуская
Беларуская (тарашкевіца)
Català
Dansk
Deutsch
Ελληνικά
English

Библиотекарь

Материал из Википедии — свободной энциклопедии

[\[править\]](#) | [править исходный текст](#)

У этого термина существуют и другие значения, см. *Библиотекарь (значения)*.

Библиотекарь — профессионал по обработке информации в библиотечной и информационной науках, который выполняет организацию и управление информационными услугами и материалами для всех, кому требуется эта информация. Как правило, библиотекари работают в государственных библиотеках или библиотеках высших или средних учебных заведений, начальных или средних школ, медиа-центрах, библиотеках в рамках предприятий или компаний. На западе некоторые библиотекари могут быть независимыми предпринимателями, работающими в качестве информационных специалистов, каталогизаторов, и по прочим специализированным направлениям.

В конце XIX - начале XX века на страницах Энциклопедического словаря Брокгауза и Ефрона было дано следующее описание этой профессии:

« Библиотекарь — звание, которое встречается уже в древнейшие времена; мы видим библиотекарей в Египте, при дворе Птолемея; в настоящее время Б. называется обыкновенно лицо, которому поручается заведование книгохранилищем или какой-либо его частью; в некоторых странах, однако, как, напр., в Германии, Б. называется исключительно то лицо, которому поручено управление всей библиотекой; лица же, заведующие отделениями библиотеки, носят разные другие названия: в Германии — Custos, Assistent, во Франции — Conservateur и т. д. С развитием библиотечного дела мало-помалу определилась та сумма качеств и знаний, которая необходима библиотекарю для успешного ведения дела; возникла наука библиотековедения, изучить которую в настоящее время можно только практически — службой в библиотеке же. От библиотекаря требуется способность к усидчивому труду, любовь к порядку, он должен знать новые языки и быть энциклопедически образованным человеком; необходимо также знание всего, что касается истории книги, книгопечатания и т. п. »^[1]



Библиотекарь, Арчимбольдо, Джузеппе, 1556

См. также [\[править\]](#) | [править исходный текст](#)

- Архивариус
- Избач

Примечания [\[править\]](#) | [править исходный текст](#)

- ↑ Библиотекарь // Энциклопедический словарь Брокгауза и Ефрона: В 86 томах (82 т. и 4 доп.). — СПб., 1890—1907.

