

QA-системы

Классическая постановка задачи

Вопросно-ответная система (QA-система) отвечает на «вопросы по тексту» на естественном языке.
Под «текстом» понимается некоторая база данных.

[WolframAlpha](#) может всё

Попробуйте ввести любой вопрос:

- Where...
- How...
- Who...
- ...

Siri интегрирована с WolframAlpha

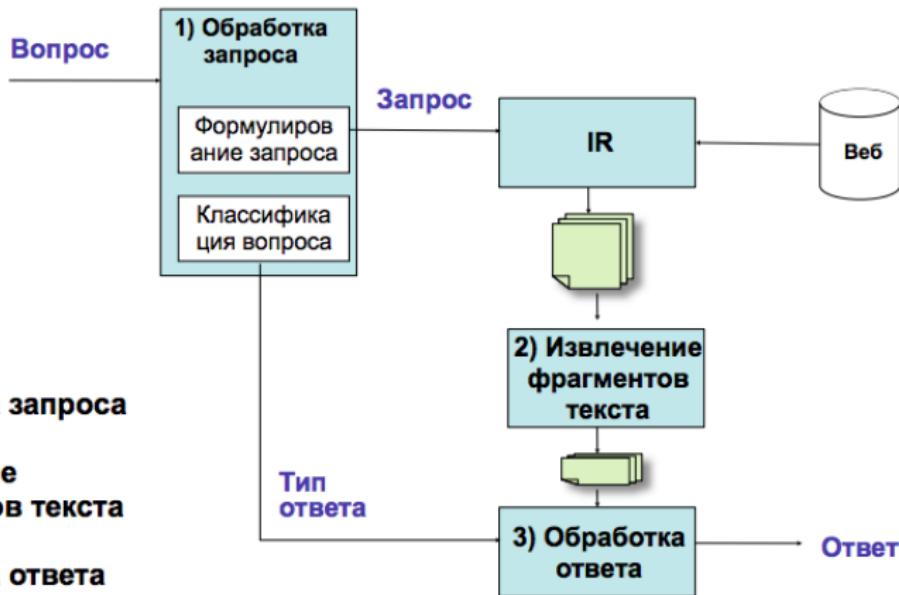
- Факт
 - *Как звали отца Татьяны Лариной?*
 - *Где растет бамбук?*
 - ...
- Список ← несколько фактов
 - *Какие страны входят в ЕС?*
 - *Какие фильмы снял Михалков?*
 - ...
- Определение ← рефериование
 - *Кто такой Марк Аврелий?*
 - *Что такое квадрокоптер?*
 - ...

Формулировки в тесте могут не совпадать с формулировкой вопроса

Подтипы фактов:

Вопрос	Основа вопроса	Тип ответа
Q555: What was the name of Titanic's captain?	What	Person
Q654: What U.S. Government agency registers trademarks?	What	Organization
Q162: What is the capital of Kosovo?	What	City
Q661: How much does one ton of cement cost?	How much	Quantity

Архитектура QA-системы



Три этапа:

1. Обработка запроса
2. Извлечение фрагментов текста
3. Обработка ответа

1. Обработка запроса

Из вопроса на естественном языке извлекаем:

- **ключевые слова** для запроса к информационно-поисковой системе
(Формулирование запроса)
- **тип ответа**, специфицирующий класс сущности, возвращаемой в качестве ответа
(Классификация вопроса)

Формулирование запроса

- Извлечь ключевые слова из вопроса

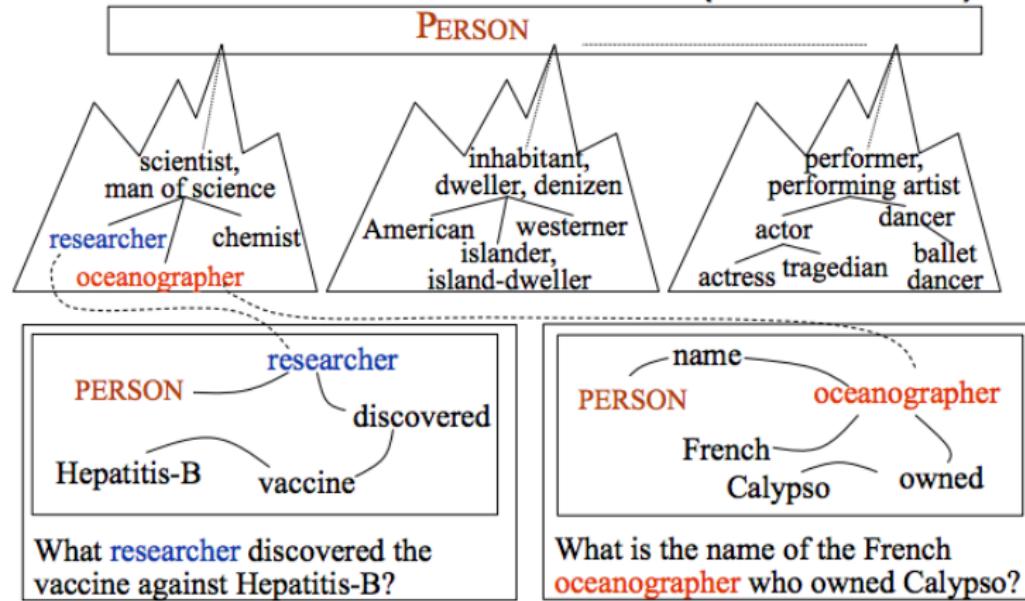
Question (from TREC QA track)	Lexical terms
Q002: What was the monetary value of the Nobel Peace Prize in 1989?	monetary, value, Nobel, Peace, Prize, 1989
Q003: What does the Peugeot company manufacture?	Peugeot, company, manufacture
Q004: How much did Mercury spend on advertising in 1993?	Mercury, spend, advertising, 1993
Q005: What is the name of the managing director of Apricot Computer?	name, managing, director, Apricot, Computer

- Расширить вопрос лексически/семантически близкими словами
- Переформулировать вопрос:
Когда родился Ленин? → Ленин родился

Классификация вопросов

Обучить классификатор по размеченному корпусу или использовать иерархию:

Таксономия типов ответов (из Wordnet)



2. Извлечение фрагментов текста

Считаем, что задача информационного поиска уже решена: IR-система вернула документы, подходящие под сформулированный (обработанный) запрос.

Нужно выбрать фрагмент (абзац/предложение), содержащее ответ:

- ① Отсеять фрагменты, не содержащие ответ
(распознавание именованных сущностей и классификация ответов)
- ② Отранжировать оставшиеся фрагменты
(правила, составленные вручную или машинное обучение)

Признаки для ранжирования

- Число именованных сущностей правильного типа в фрагменте
- Число ключевых слов из вопроса в фрагменте
- Наиболее длинная последовательность ключевых слов запроса в фрагменте
- Ранг документа (IR), содержащего фрагмент
- Плотность ключевых слов из вопроса в фрагменте
- Пересечение N-грамм вопроса и фрагмента

Извлечение фрагментов из Веба

Можно воспользоваться сниппетами от IR-систем.

Яндекс

что такое сниппет

ПОИСК КАРТИНКИ ВИДЕО КАРТЫ МАРКЕТ НОВОСТИ ПЕРЕВОДЧИК ЕЩЁ

W Сниппет — Википедия
ru.wikipedia.org > Сниппет

Сниппет (англ. **snippet** — отрывок, фрагмент) — фрагмент исходного текста или кода программы, применяемый в: поисковых системах, текстовых редакторах и средах разработки.

Я Сниппеты страниц

yandex.ru > Помощь > Вебмастер > .../site-description.xml

Что такое сниппет? Сниппет — это блок информации о найденном документе, который отображается в результатах поиска.

3. Обработка ответа

Имеем отранжированный список подходящих фрагментов.

Классы алгоритмов для получения ответа:

- составление ответа по шаблону
- составление ответа из N-грамм

Составление ответа по шаблону

На основе регулярных выражений:

- по информации о типе ответа

Если тип ответа «person», извлечь из фрагмента именованные сущности «person»

- по шаблонам типа *such as*

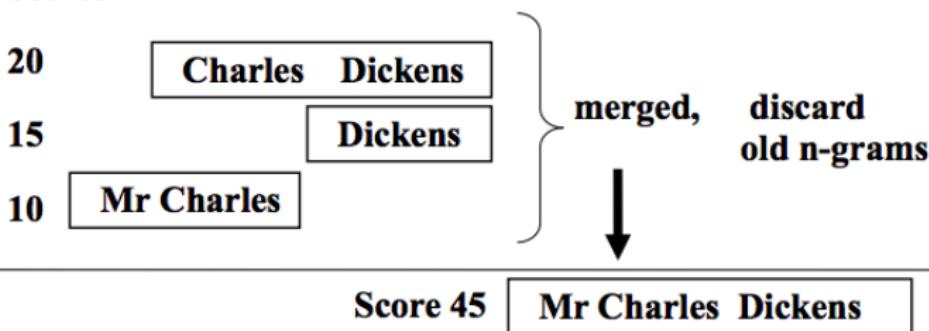
Pattern	Question	Answer
<AP> such as <QP>	<i>What is autism?</i>	<i>developmental disorders such as autism</i>

Составление ответа из N-грамм

- 1 Назначить N-граммам вес, пропорциональный количеству сниппетов, в которых они встретились (учитывать тип ответа).
- 2 Собрать ответ по пересечениям топа N-грамм

Пример: «Who created the character of Scrooge?»

Scores



Вспоминаем типы ответов

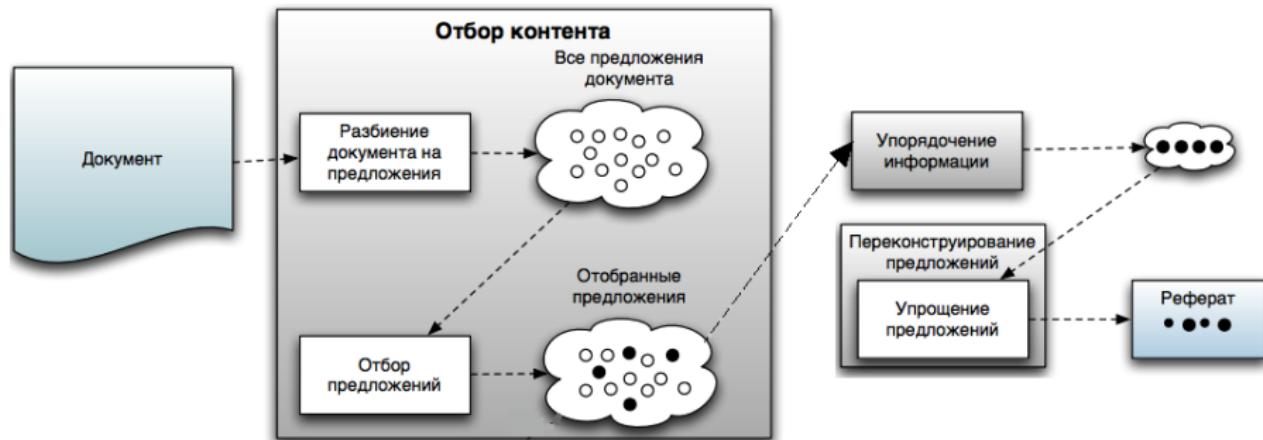
- Факт
- Список ← несколько фактов
- Определение ← рефериование
 - *Кто такой Марк Аврелий?*
 - *Что такое квадрокоптер?*
 - ...

Подходы к суммаризации текстов

На примере Геттисбергской речи Авраама Линкольна

- Extraction-based summarization (реферирование)
 - составление выдержек (выделение предложений):
Восемьдесят семь лет назад наши отцы ступили на эту землю, чтобы создать новую нацию
- Abstraction-based summarization (аннотирование)
 - генерация нового текста
(не обязательно теми же словами):
В этой речи Авраам Линкольн призывает вспомнить солдат, которые отдали свои жизни в битве при Геттисберге

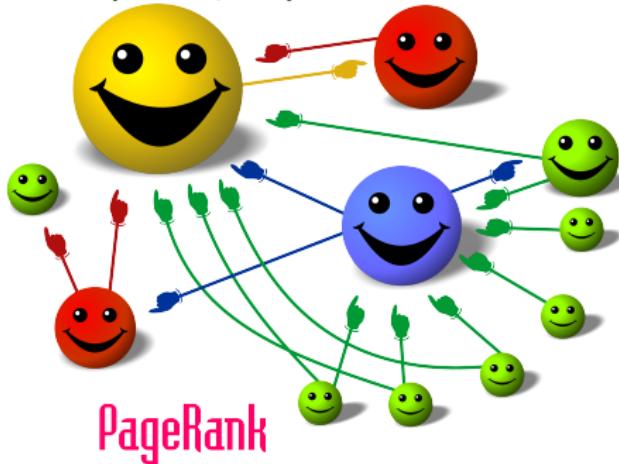
Архитектура системы автореферирования



1. Отбор контента

Выбор предложений

- по ключевым словам (tf-idf, log likelihood ratio, ...)
 - по мерам центральности



- по признакам: позиция, обобщающие фразы (*to sum up*), длина предложения, связность.

2. Упорядочение и 3. Упрощение

Для одного документа:

- остается исходный порядок

Для нескольких документов:

- кластеризация по содержанию

Упрощение может включать удаление избыточных второстепенных членов.

Stanford Question Answering Dataset

- Содержит текстовые фрагменты - абзацы из «высококачественных» статей Википедии.
- Каждый фрагмент сопровождается вопросами на «понимание прочитанного».
- На каждый вопрос записан один или несколько ответов: отрывков = сегментов = спансов (spans) текста.

Пример тройки SQuAD

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, *Il milione* (or, *The Million*, known in English as the *Travels of Marco Polo*), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge **through contact with Persian traders** since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: **through contact with Persian traders**

Paragraph 1 of 43

Spend around 4 minutes on the following paragraph to ask 5 questions! If you can't ask 5 questions, ask 4 or 3 (worse), but do your best to ask 5. Select the answer from the paragraph by clicking on 'Select Answer', and then highlight the smallest segment of the paragraph that answers the question.

Oxygen is a chemical element with symbol O and atomic number 8. It is a member of the chalcogen group on the periodic table and is a highly reactive nonmetal and oxidizing agent that readily forms compounds (notably oxides) with most elements. By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O₂.

2. Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere. However, monitoring of atmospheric oxygen levels show a global downward trend, because of fossil-fuel burning. Oxygen is the most abundant element by mass in the Earth's crust as part of oxide compounds such as silicon dioxide, making up almost half of the crust's mass.

When asking questions, **avoid using** the same words/phrases as in the paragraph. Also, you are encouraged to pose **hard questions**.

Ask a question here. Try using your own words

Select Answer

<https://rajpurkar.github.io/mlx/qa-and-squad/>

Понять, что «referred to as» = «call».

Passage Segment

...The Rankine cycle is sometimes referred to as a practical Carnot cycle...

Question

What is the Rankine cycle sometimes called?

Понять, что «the European Parliament» и «the Council of the European Union» являются «government bodies», хотя этой информации нет в данном текстовом фрагменте.

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

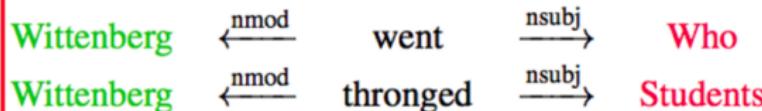
Which governing bodies have veto power?

Сложности вопросов

Без синтаксических вариаций всё хорошо: синтаксическая структура дает всю информацию, необходимую для ответа.

...**Students** thronged to Wittenberg to hear Luther speak....

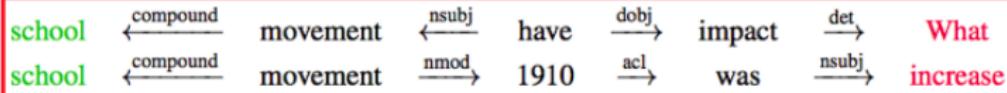
Who went to Wittenberg to hear Luther speak?



Структура деревьев разбора вопроса и предложения в текстовом фрагменте может сильно различаться.

...During the mass high school education movement from 1910 – 1940 , there was an increase in skilled workers...

What impact did the high school education movement have on the presence of skilled workers?



Несколько предложений + кореферентные связи

Passage Segment

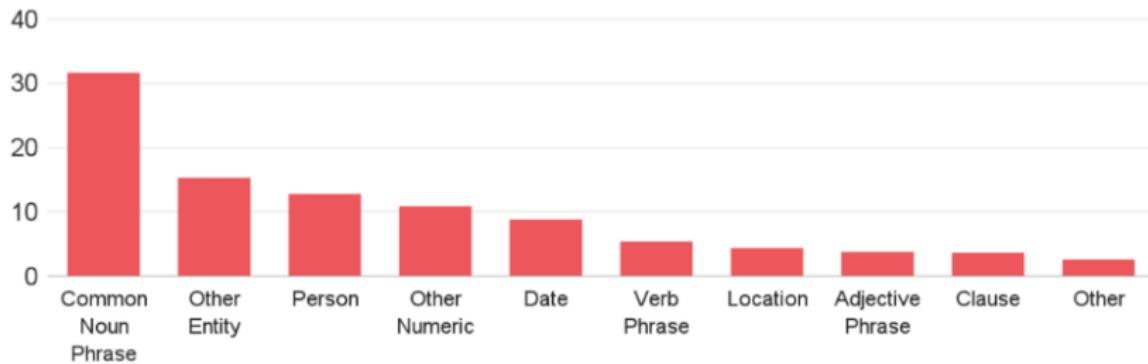
...The V&A Theatre and Performance galleries opened in March 2009. ...
They hold the UK's biggest national collection of material about live performance...

Question

What collection does the V&A
Theatre & Performance galleries
hold?

Разнообразие ответов

Многие QA-системы используют ожидаемый тип ответа при ответе на вопрос. Например, если есть вопрос «сколько», то рассматриваются только числовые ответы-кандидаты. В SQuAD ответы не ограничены определенным типом, могут включать не-сущности и длинные фразы.



SQuAD использует два разных показателя:

- **Exact Match metric:** измеряет процент предсказанных ответов, точно соответствующих одному из истинных ответов.
- **F1 score metric:** измеряет среднее совпадение между предсказанным ответом и истинным ответом.

Оценка Human Performance

Один из ответов рассматривается как предсказанный, остальные - истинные.

- Exact Match metric: 82,3
- F1 score metric: 91,2

**Human Performance
(91.2 F1)**



- a **sliding window baseline**: ранжируем всех возможных кандидатов, сопоставляя их с вопросом в форме BoW.
F1 score metric: 20
- **logistic regression baseline**: использует различные признаки, в частности **lexicalized features** и **dependency tree path features**.

Lexicalized Features

Леммы вопроса соединяются в пары с леммами ответа

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

Question

What causes precipitation to fall?

Answer Candidate

gravity

- Between question and answer

cause---gravity

precipitation---gravity

fall---gravity

what---gravity

Леммы вопроса соединяются в пары с леммами слов рядом с ответом

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

Question

What causes precipitation to fall?

Answer Candidate

gravity

- Between question and passage sentence (around candidate)

cause---under

cause---fall

precipitation---fall

fall---under

Dependency Features

В нужном предложении из фрагмента текста строится
путь зависимостей от слов, которые встречаются в
вопросе, до ответа.

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

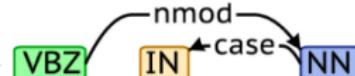
Question

What causes precipitation to fall?

Answer Candidate

gravity

- Path from passage sentence words (that also occur in question) to answer



- Combined with path from wh-word to question word.



Logistic Regression Baseline

Модель правильно выбирает предложение, содержащее ответ, с точностью 79,3%. Основная сложность - определение точной длины.



Практически любую задачу NLP можно представить как задачу QA

I: Mary walked to the bathroom.

I: Sandra went to the garden.

I: Daniel went back to the garden.

I: Sandra took the milk there.

Q: Where is the milk?

A: garden

I: Everybody is happy.

Q: What's the sentiment?

A: positive

I: Jane has a baby in Dresden.

Q: What are the named entities?

A: Jane - person, Dresden - location

I: Jane has a baby in Dresden.

Q: What are the POS tags?

A: NNP VBZ DT NN IN NNP .

I: I think this model is incredible

Q: In French?

A: Je pense que ce modèle est incroyable

Препятствия к созданию универсальной QA-системы

1. Разные архитектуры для разных задач

Task	State of the art model
Question answering (babl)	Strongly Supervised MemNN (Weston et al 2015)
Sentiment Analysis (SST)	Tree-LSTMs (Tai et al. 2015)
Part of speech tagging (PTB-WSJ)	Bi-directional LSTM-CRF (Huang et al. 2015)

2. Сложное многозадачное обучение несвязанных задач

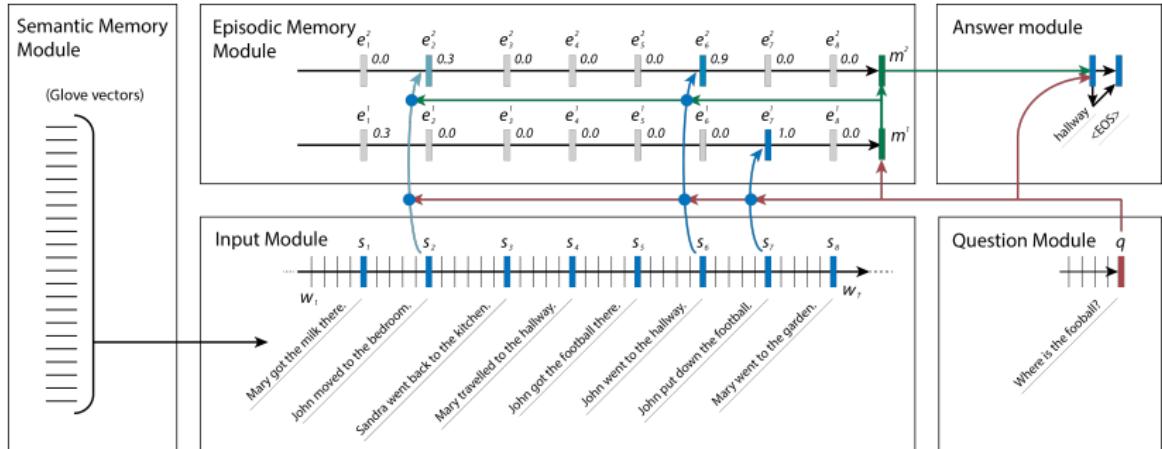
Архитектура для любой QA-задачи.

«Imagine having to read an article, memorize it, then get asked various questions → **Hard!**

You can't store everything in working memory

Optimal: give you the input data, give you the question, allow as many glances as possible».

Архитектура динамической нейронной сети



Подробности: [статья]

Implementation of Improved Dynamic Memory Networks: <http://ethancaballero.pythonanywhere.com/>

Task	MemNN	DMN	Task	MemNN	DMN
1: Single Supporting Fact	100	100	11: Basic Coreference	100	99.9
2: Two Supporting Facts	100	98.2	12: Conjunction	100	100
3: Three Supporting facts	100	95.2	13: Compound Coreference	100	99.8
4: Two Argument Relations	100	100	14: Time Reasoning	99	100
5: Three Argument Relations	98	99.3	15: Basic Deduction	100	100
6: Yes/No Questions	100	100	16: Basic Induction	100	99.4
7: Counting	85	96.9	17: Positional Reasoning	65	59.6
8: Lists/Sets	91	96.5	18: Size Reasoning	95	95.3
9: Simple Negation	100	100	19: Path Finding	36	34.5
10: Indefinite Knowledge	98	97.5	20: Agent's Motivations	100	100
Mean Accuracy (%)				93.3	93.6