

**«Программы-  
переводчики.**

**Возможности систем  
распознавания текстов»**

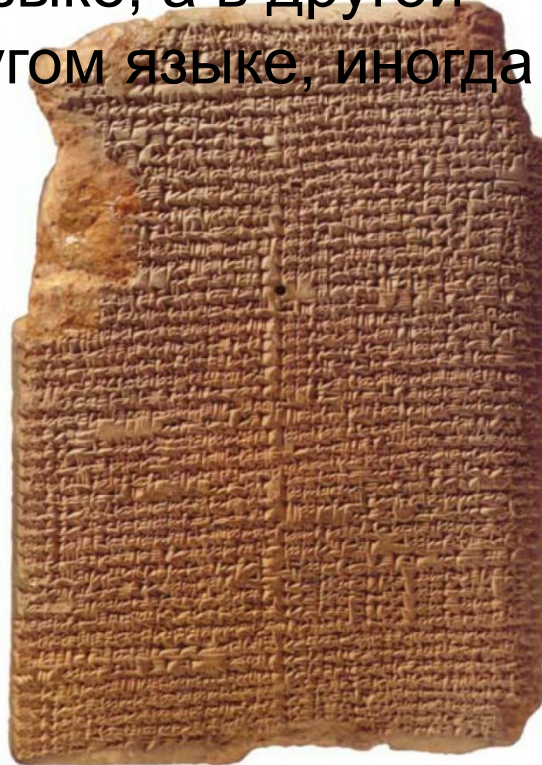
# Основные понятия темы :

- 1. Компьютерные словари
- 2. Машины-переводчики
- 3. Системы оптического распознавания текста.

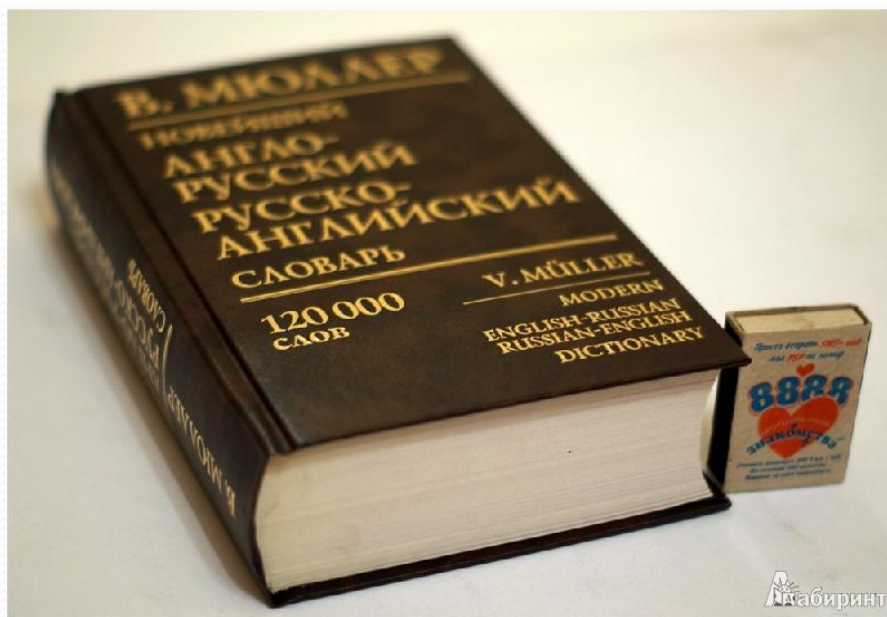
## Компьютерные словари.

Словари необходимы для перевода текстов с одного языка на другой.

Первые словари были созданы около 5 тысяч лет назад в Шумере и представляли собой глиняные таблички, разделенные на две части. В одной части записывалось слово на шумерском языке, а в другой — аналогичное по значению слово на другом языке, иногда с краткими пояснениями.



Современные словари построены по такому же принципу. В настоящее время существуют тысячи словарей для перевода между сотнями языков (англо-русский, немецко-французский и так далее), причем каждый из них может содержать десятки тысяч слов. В бумажном варианте словарь представляет собой толстую книгу объемом в сотни страниц, где поиск нужного слова является достаточно трудоемким процессом.



А  
Б  
В  
Г  
Д  
Е  
Е  
Ж  
З  
И  
Й  
К  
Л  
М  
Н  
О  
П  
Р  
С  
Т  
У  
Ф  
Х  
Ц  
Ч  
Ш  
Щ  
Ъ  
Ы  
Ь  
Э  
Ю

пеликан	['pelɪkən]	(пѐликѐн)	pelican		
пенал	['pensɪlbɒks]	(пѐнсѝлбѝкс)	pencil-box		
пень	[stʌmp]	(стѝмп)	stump		
первый	[fɜːst]	(фѐст)	first		
первый этаж	['fɜːstflɔː]	(фѐст флѝо)	first floor		
перекрёсток	['krɒsrəʊdz]	(крѝсрѝоудѝз)	crossroads		
перец	['pepə]	(пѝпѝ)	pepper		
перечница	['pepə.kæstə]	(пѝпѝ кѝстѝ)	pepper-caster		
перила	['bænistəz]	(бѝнѝстѝз)	banisters		
перина	['feðəbed]	(фѐдѝбѝд)	feather-bed		
перо	['feðə]	(фѝдѝ)	feather		
персик	['pi:tʃ]	(пѝч)	peach		
перчатки	[glɒvz]	(глѝвѝз)	gloves		
песок	['sænd]	(сѝнд)	sand		
песочница	['sændpɪt]	(сѝндпѝт)	sandpit		
песочные часы	['aʊəglɑːs]	(ѝуѝглѝс)	hour-glass		
петрушка	['pɑːslɪ]	(пѝслѝ)	parsley		
петуния	['piːtjuːnjə]	(пѝтѝнѝнѝ)	petunia		
петух	['rʊstə]	(рѝстѝ)	rooster		
печенье	['bɪskɪts]	(бѝскѝтѝ)	biscuits		

пешеходный переход	[piːdestriənˈkrɒsɪŋ]	(пѝдѝстѝриѝн крѝсѝнѝ)	pedestrian crossing		
пианино	['piːnɪnoʊ]	(пѝнѝноу)	piano		
пиво	[biːə]	(бѝѝ)	beer		
пижама	[pɪˈdʒæməz]	(пѝдѝжѝмѝз)	pyjamas		
пила	[sɔː]	(сѝ)	saw		
пингвин	['pɛŋɡwɪn]	(пѝнѝгѝн)	penguin		
пион	['piːni]	(пѝнѝнѝ)	peony		
пипетка	[piːpet]	(пѝпѝт)	pipette		
пирог	[paɪ]	(пѝѝ)	pie		
пирожное	['peɪstrɪ]	(пѝѝстѝрѝ)	pastry		
писатель	['raɪtə]	(рѝѝтѝ)	writer		
письменный стол	['desk]	(дѝск)	desk		
пицца	['pɪtsə]	(пѝтѝсѝ)	pizza		
плавание	['swɪmɪŋ]	(сѝѝмѝнѝнѝ)	swimming		
плавки	['swɪmɪŋˈtrʌŋks]	(сѝѝмѝнѝнѝ трѝнѝкс)	swimming trunks		
плавник	[fɪn]	(фѝн)	fin		
пластилин	['plæstɪsɪn]	(плѝстѝсѝнѝ)	plasticine		
пластинка	['rekɔːd]	(рѝкѝд)	record		
пластырь	['plæstə]	(плѝстѝ)	plaster		
платье	[dres]	(дрѝс)	dress		

А  
Б  
В  
Г  
Д  
Е  
Е  
Ж  
З  
И  
Й  
К  
Л  
М  
Н  
О  
П  
Р  
С  
Т  
У  
Ф  
Х  
Ц  
Ч  
Ш  
Щ  
Ъ  
Ы  
Ь  
Э  
Ю

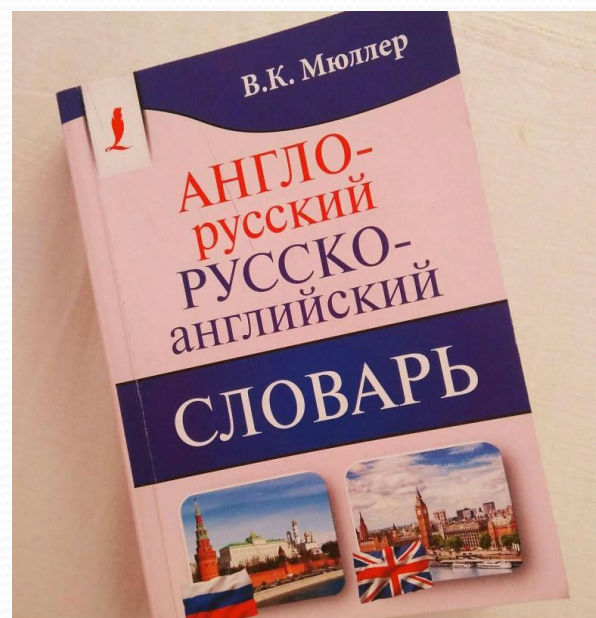
Пример словаря для детей

Компьютерные словари могут содержать переводы на разные языки сотен тысяч слов и словосочетаний, а также предоставляют пользователю дополнительные возможности.

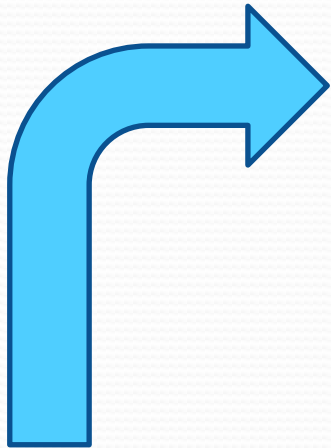
Во-первых, компьютерные словари могут являться многоязычными, так как дают пользователю возможность выбрать языки и направление перевода (например, англо-русский, испано-русский и так далее).



© WILDBERRIES



Во-вторых, компьютерные словари могут кроме основного словаря общеупотребительных слов содержать десятки специализированных словарей по областям знаний (техника, медицина, информатика и др.).

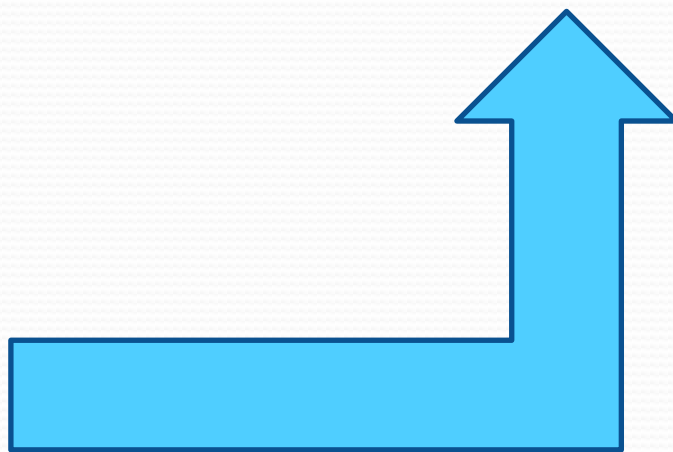


АНГЛО-РУССКИЙ  
МЕДИЦИНСКИЙ  
ЭНЦИКЛОПЕДИЧЕСКИЙ  
СЛОВАРЬ

---

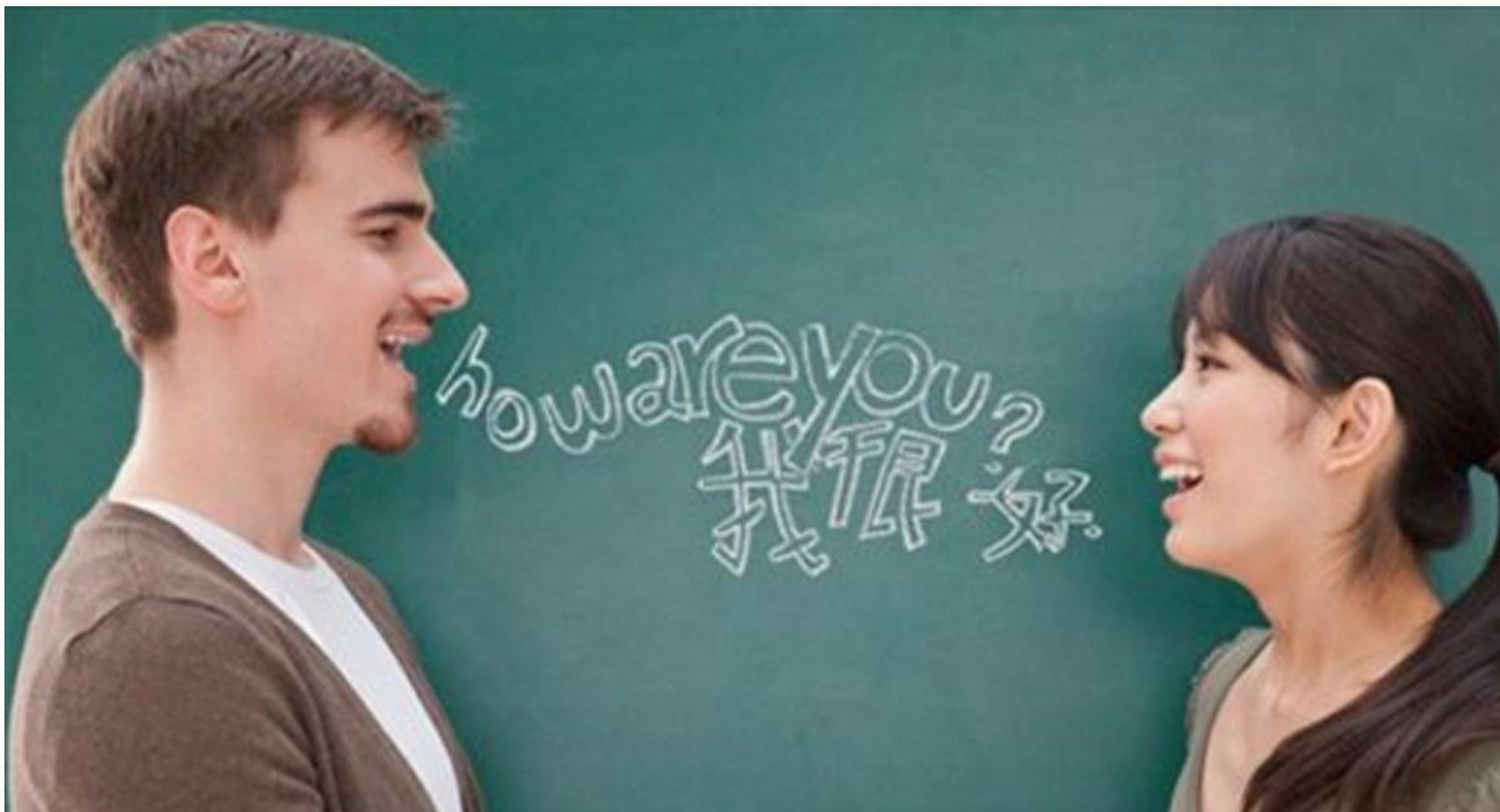
*Stedman's*  
MEDICAL  
DICTIONARY

В-третьих, компьютерные словари обеспечивают быстрый поиск словарных статей: «быстрый набор», когда в процессе набора слова возникает список похожих слов; доступ к часто используемым словам по закладкам; возможность ввода словосочетаний и др.





В-четвертых, компьютерные словари могут являться мультимедийными, то есть предоставлять пользователю возможность прослушивания слов в исполнении дикторов, носителей языка.



## Системы машинного перевода.

Происходящая в настоящее время глобализация нашего мира приводит к необходимости обмена документами между людьми и организациями, находящимися в разных странах мира и говорящими на различных языках. В этих условиях использование традиционной технологии перевода «вручную» тормозит развитие межнациональных контактов. Перевод многостраничной документации вручную требует длительного времени и высокой оплаты труда переводчиков. Перевод полученного по электронной почте письма или просматриваемой в браузере Web-страницы необходимо осуществить немедленно, и нет возможности и времени пригласить переводчика.

Системы машинного перевода позволяют решить эти проблемы.

Они, с одной стороны, способны переводить многостраничные документы с высокой скоростью (одна страница в секунду) и, с другой стороны, переводить Web-страницы «на лету», в режиме реального времени.

Лучшими среди российских систем машинного перевода считаются PROMT и «Сократ».

Системы машинного перевода осуществляют перевод текстов, основываясь на формальном «знании» языка (синтаксиса языка — правил построения предложений, правил словообразования) и использовании словарей.

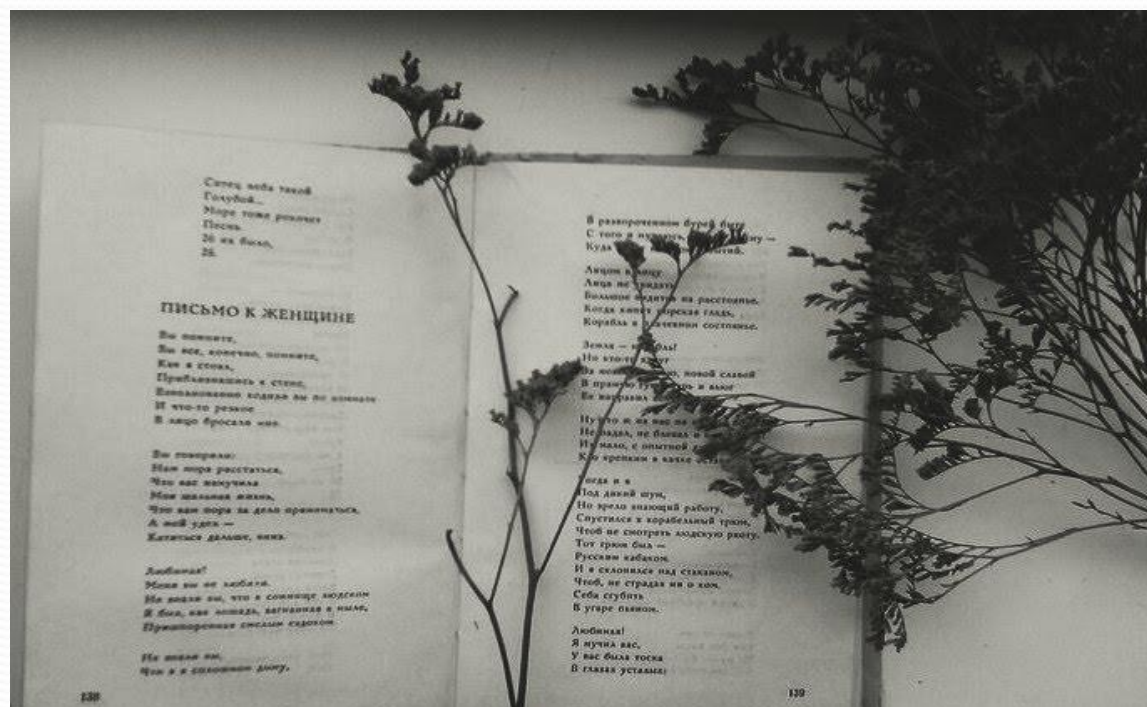
Программа-переводчик сначала анализирует текст на одном языке, а затем конструирует этот текст на другом языке.

Лучшими среди российских систем машинного перевода считаются PROMT и «Сократ».



Современные системы машинного перевода позволяют достаточно качественно переводить техническую документацию, деловую переписку и другие специализированные тексты.

Однако они неприменимы для перевода художественных произведений, так как не способны адекватно переводить метафоры, аллегории и другие элементы художественного творчества человека.



## Достоинства программ-переводчиков

- наличие словарей по специальностям, мгновенный поиск, расположение в несколько окон, возможность одновременного обзора нескольких вариантов перевода;
- возможность создания собственного словаря пользователя
- подключение к текстовому редактору Microsoft Office, что позволяет, не отрываясь от оригинала и его перевода, выбирать подходящие соответствия для перевода;
- возможность перевода с разных языков;
- компактность, наличие практически неограниченного объема информации в компьютере;

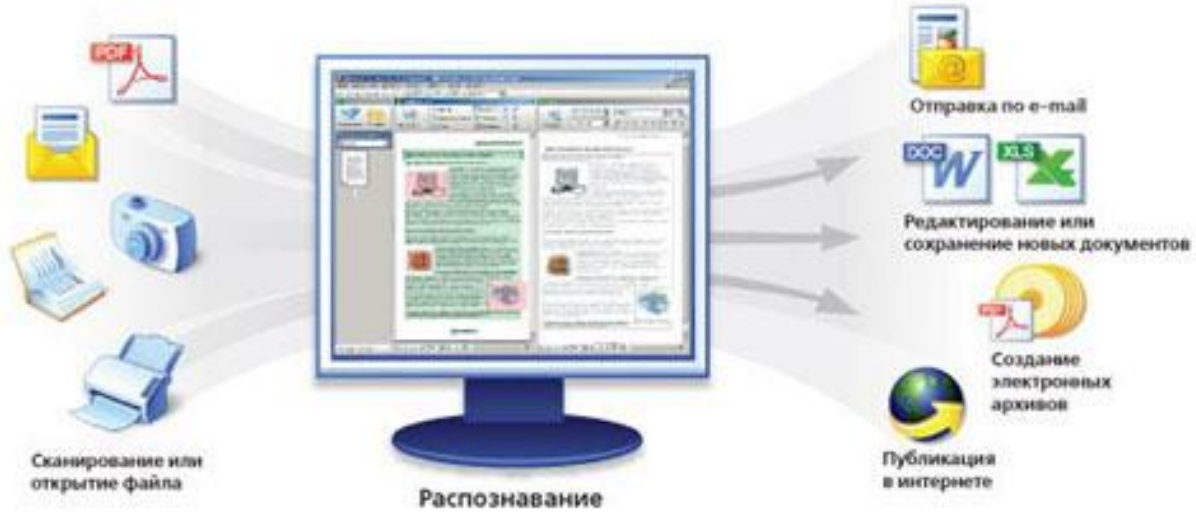
## Недостатки программ-переводчиков

- словари-переводчики не соблюдают правила грамматики, стилистики и лексики, не учитывают игру слов, художественные приемы;
- чаще всего текстовые переводчики выбирают одно из значений многозначного слова, которое может не соответствовать контексту;
- при отсутствии слова в словаре не переводят его;
- электронные словари-переводчики часто выдают несколько вариантов перевода слова на другой язык, не объясняя разницы в тех или иных словарных соответствиях, что затрудняет правильный выбор того или иного соответствия в данном контексте.

**Системы оптического распознавания символов** используются при создании электронных библиотек и архивов путем перевода книг и документов в цифровой компьютерный формат.

После обработки документа сканером получается графическое изображение документа (графический образ). Но графический образ еще не является текстовым документом. Человеку достаточно взглянуть на лист бумаги с текстом, чтобы понять, что на нем написано. С точки зрения компьютера, документ после сканирования превращается в набор разноцветных точек, а вовсе не в текстовый документ.

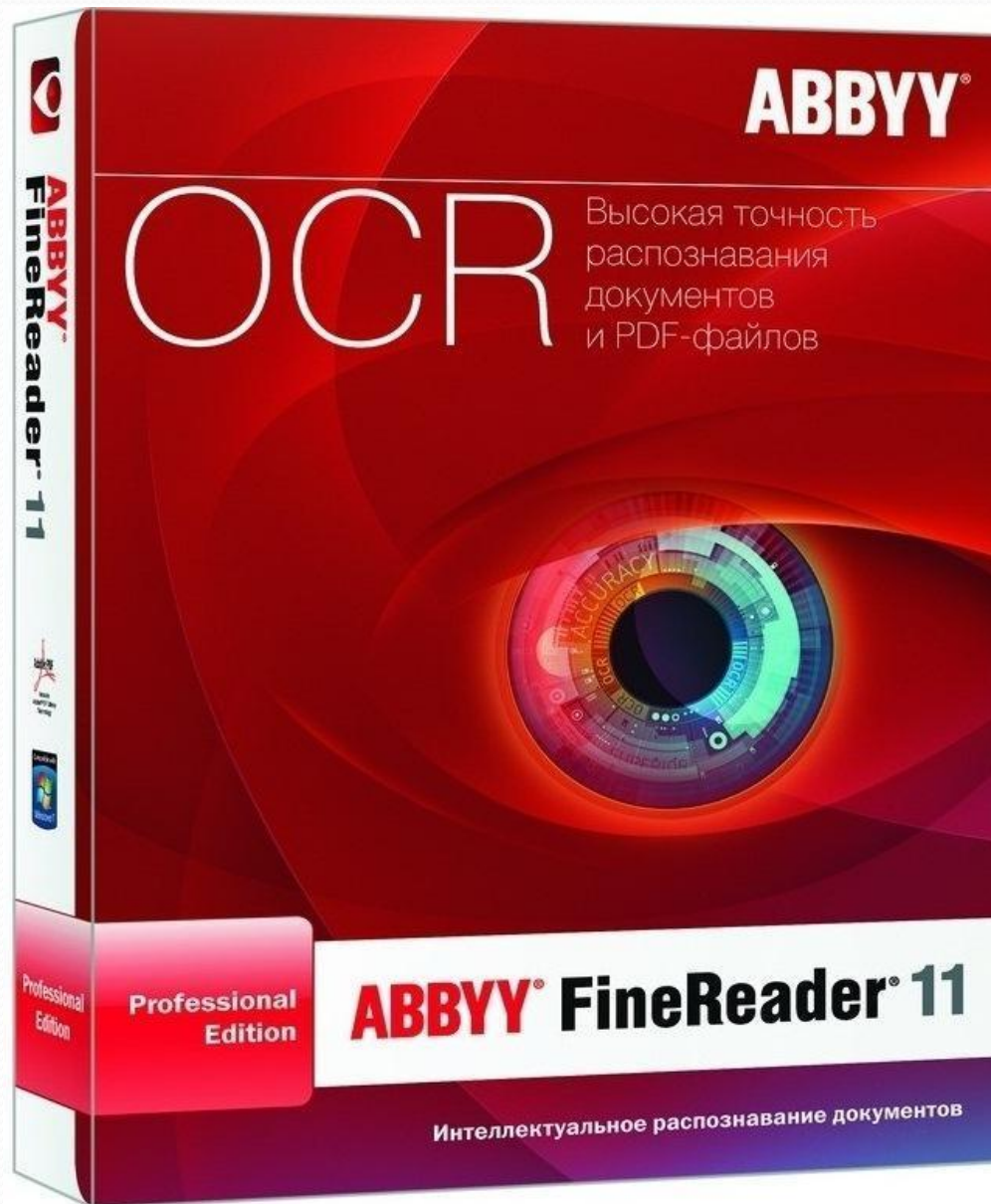
Проблема распознавания текста в составе точечного графического изображения является весьма сложной. Подобные задачи решают с помощью специальных программных средств, называемых средствами распознавания образов. Реальный технический прорыв в этой области произошел лишь в последние годы. До этого распознавание текста было возможно только путем сравнения обнаруженных конфигураций точек со стандартным образцом (эталонном, хранящимся в памяти компьютера). Авторы программ задавали критерий «похожести», используемый при идентификации символов.





Программа **FineReader** выпускается отечественной компанией ABBYY Software ([www.bitsoft.ru](http://www.bitsoft.ru)). Эта программа предназначена для распознавания текстов на русском, английском, немецком, украинском, французском и многих других языках, а также для распознавания смешанных текстов.

Она позволяет объединять сканирование и распознавание в одну операцию, работать с пакетами документов и с бланками. Программу можно обучать для повышения качества распознавания неудачно напечатанных текстов или сложных шрифтов. Она позволяет редактировать распознанный текст и проверять его орфографию.



**ABBYY®**

OCR

Высокая точность  
распознавания  
документов  
и PDF-файлов



Professional  
Edition

Professional  
Edition

**ABBYY® FineReader® 11**

Интеллектуальное распознавание документов

Инструменты распознавания текстов и компьютерного перевода -  
Информатика 7 класс #27 - Инфоурок.mp4