



Лекция №8

Кросс-платформенное программирование



Виды диаграмм

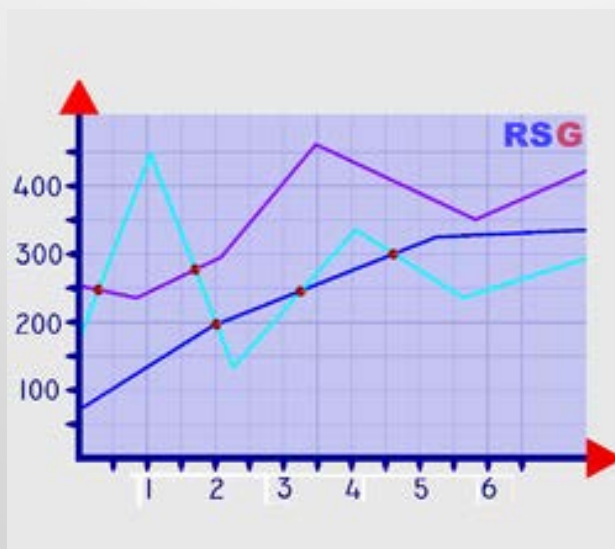
Регулярные выражения

Диаграммы

- Диагра́мма (греч. Διάγραμμα (diagramma) — изображение, рисунок, чертёж) — графическое представление данных, позволяющее быстро оценить соотношение нескольких величин. Представляет собой геометрическое символическое изображение информации с применением различных приёмов техники визуализации.

Основные типы диаграмм

Диаграммы-линии или графики — это тип диаграмм, на которых полученные данные изображаются в виде точек, соединённых прямыми линиями. Точки могут быть как видимыми, так и невидимыми (ломанные линии). Также могут изображаться точки без линий (точечные диаграммы). Для построения диаграмм-линий применяют прямоугольную систему координат.



Диаграммы-области

- Диаграммы-области — это тип диаграмм, схожий с линейными диаграммами способом построения кривых линий. Отличается от них тем, что область под каждым графиком заполняется индивидуальным цветом или оттенком. Преимущество данного метода в том, что он позволяет оценивать вклад каждого элемента в рассматриваемый процесс.

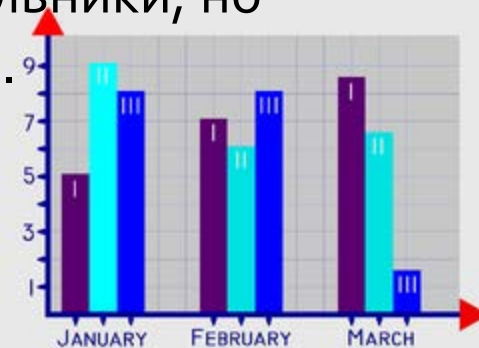


Столбчатые и линейные диаграммы (гистограммы)

Классическими диаграммами являются столбчатые и линейные (полосовые) диаграммы. Также они называются гистограммами. Столбчатые диаграммы в основном используются для наглядного сравнения полученных статистических данных или для анализа их изменения за определённый промежуток времени. Построение столбчатой диаграммы заключается в изображении статистических данных в виде вертикальных прямоугольников или трёхмерных прямоугольных столбиков. Каждый столбик изображает величину уровня данного статистического ряда. Все сравниваемые показатели выражены одной единицей измерения, поэтому удаётся сравнить статистические показатели данного процесса.

Разновидностями столбчатых диаграмм являются линейные (полосовые) диаграммы. Они отличаются горизонтальным расположением столбиков. Столбчатые и линейные диаграммы взаимозаменяемы, рассматриваемые в них статистические показатели могут быть представлены как вертикальными, так и горизонтальными столбиками. В обоих случаях для изображения величины явления используется одно измерение каждого прямоугольника — высота или длина столбика. Поэтому и сфера применения этих двух диаграмм в основном одинакова.

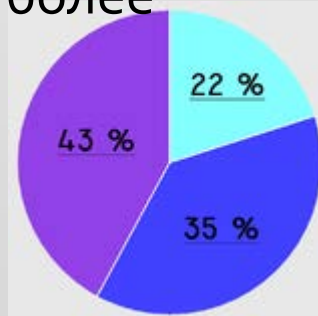
Столбчатые диаграммы могут изображаться и группами (одновременно расположенными на одной горизонтальной оси с разной размерностью варьирующих признаков). Образующие поверхности столбчатых и линейных диаграмм могут представлять собой не только прямоугольники, но также квадраты, треугольники, трапеции и т. д.



Круговые (секторные) диаграммы

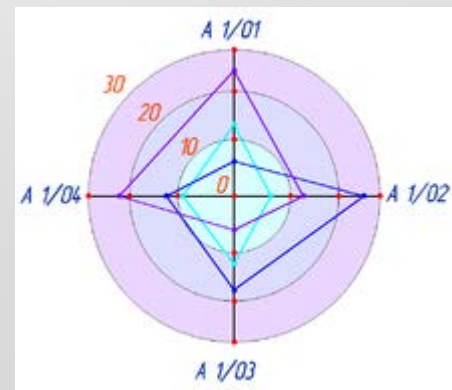
Достаточно распространённым способом графического изображения структуры статистических совокупностей является секторная диаграмма, так как идея целого очень наглядно выражается кругом, который представляет всю совокупность. Относительная величина каждого значения изображается в виде сектора круга, площадь которого соответствует вкладу этого значения в сумму значений. Этот вид графиков удобно использовать, когда нужно показать долю каждой величины в общем объёме. Сектора могут изображаться как в общем круге, так и отдельно, расположенными на небольшом удалении друг от друга.

Круговая диаграмма сохраняет наглядность только в том случае, если количество частей совокупности диаграммы небольшое. Если частей диаграммы слишком много, её применение неэффективно по причине несущественного различия сравниваемых структур. Недостаток круговых диаграмм — малая ёмкость, невозможность отразить более широкий объём полезной информации



Круговые (секторные) диаграммы

- В отличие от линейных диаграмм, в радиальных или сетчатых диаграммах более двух осей. По каждой из них производится отсчёт от начала координат, находящегося в центре. Для каждого типа полученных значений создаётся своя собственная ось, которая исходит из центра диаграммы. Радиальные диаграммы напоминают сетку или паутину, поэтому иногда их называют сетчатыми. Преимущество радиальных диаграмм в том, что они позволяют отображать одновременно несколько независимых величин, которые характеризуют общее состояние структуры статистических совокупностей. Если отсчёт производить не с центра круга, а с окружности, то такая диаграмма будет называться спиральной диаграммой



Пузырьковая диаграмма

Пузырьковая диаграмма — это разновидность точечной диаграммы, в которой точки данных заменены пузырями, причем их размер служит дополнительным измерением данных. На пузырьковой диаграмме, как и на точечной, нет оси категорий — и горизонтальная, и вертикальная оси являются осями значений. В дополнение к значениям X и значениям Y , наносимым на точечную диаграмму, на пузырьковой диаграмме показаны также значения Z (размер).

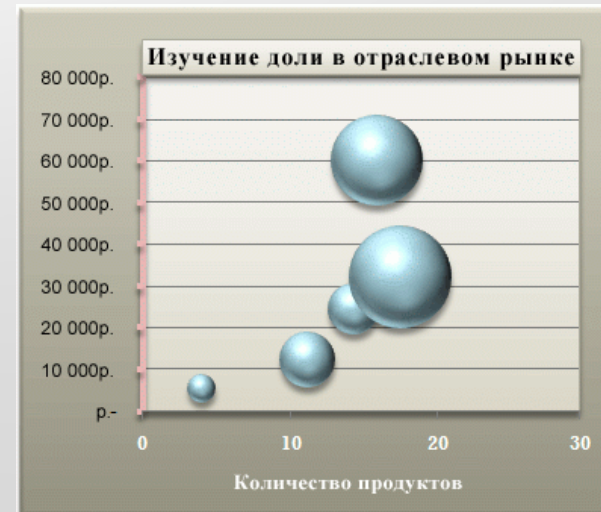
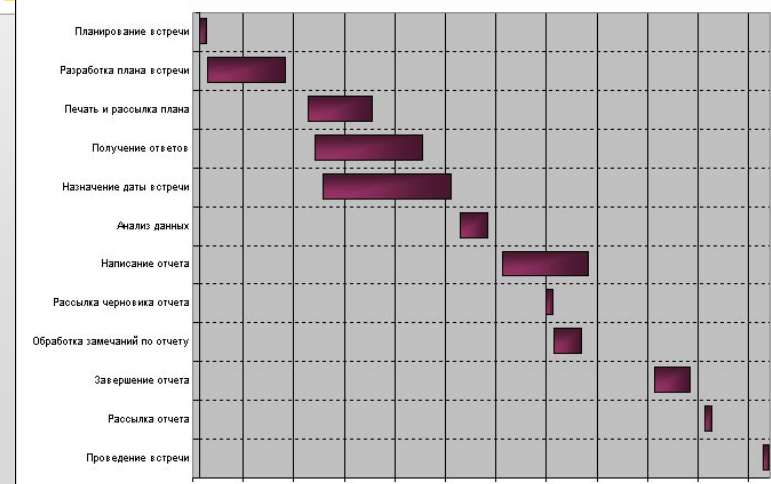



Диаграмма Ганта

Диагра́мма Га́нта (англ. Gantt chart, также ленточная диаграмма, график Ганта) — это популярный тип столбчатых диаграмм (гистограмм), который используется для иллюстрации плана, графика работ по какому-либо проекту. Является одним из методов планирования проектов.





UNIX: регулярные выражения,
POSIX/PCRE,
GREP

Регулярные выражения

Regular expressions (RegExp, RegEx):

- формальный язык поиска и осуществления манипуляций с подстроками в тексте;
- использование метасимволов (wildcard characters);
- «шаблон» («маска») - задает правило поиска.

Концепция регулярных выражений

Общая для множества программ и языков программирования:

- различные реализации отличаются в деталях, но принципы создания регулярных выражений везде одинаковы.

Базовые понятия

- Перечисление

| - разделяет допустимые варианты

"gray|grey"

- Группировка

() – определяет область действия и приоритет операторов

"gr(a|e)y"

Базовые понятия(продолжение)

- Квантификация

- $\{m, n\}$

- $\{m, \}$

- $\{, n\}$

- $? \{0, 1\}$

- $* \{0, \}$

- $+ \{1, \}$

“go+gle” соответствует gogle, google и т.д., но не ggle

Регулярные выражения в UNIX

- Синтаксис BRE (basic regular expressions) на данный момент определён POSIX как устаревший;
- широко распространён из соображений обратной совместимости;
- многие UNIX-утилиты используют такие регулярные выражения по умолчанию.

POSIX стандарт

Набор символов

- . Точка - любой символ
- [<символы>] - квадратные скобки класс символов ("любое из")
- [^<символы>] - негативный класс символов ("любое кроме")
- - обозначение последовательности в классе символов ("[0-9]" - цифры)

Квантификаторы и якоря в POSIX

Квантификаторы:

- * "звёздочка" или знак умножения {0,}
- + плюс {1,}
- ? вопросительный знак {0,1}

Якоря:

- ^ привязка к началу строки
- \$ привязка к концу строки

Расширенные регулярные выражения

К традиционному UNIX-синтаксису добавлены символы:

- +
- ?
- |

Отменен обязательный символ \

PCRE стандарт

- Perl compatible regular expressions
- i регистронезависимый поиск
- m многостроковый режим.
- "\$" совпадают с началом и концом отдельных строк.
- s символ "." (точка) совпадает и с переносом строки (по умолчанию -- нет)
- A привязка к началу текста
- E заставляет символ "\$" совпадать только с концом текста.
- U Инвертирует "жадность" для каждого квантификатора (если после квантификатора стоит "?", квантификатор перестает быть "жадным").

Операторы регулярных выражений

Оператор	Назначение
. (точка)	Любой одиночный символ.
^ (крышка)	Пустая последовательность, возникающая в начале строки.
\$ (знак доллара)	Пустая последовательность, возникающая в конце строки.
A	Буква A в верхнем регистре.
a	Буква a в нижнем регистре.
\d	Одиночная цифра.
\D	Одиночный символ, не являющийся цифрой.
\w	Любой одиночный цифро-буквенный символ, этот оператор - синоним [:alnum:].
[A-E]	Любая буква из A, B, C, D или E в верхнем регистре.
[^A-E]	Любой символ, кроме буквы из A, B, C, D или E в верхнем регистре.

Операторы регулярных выражений (продолжение)

Оператор	Назначение
$X?$	Одна или ни одной заглавной буквы X.
X^*	Ноль или более заглавной буквы X.
X^+	Одна или более заглавная буква X.
$X\{n\}$	Точно n заглавных букв X.
$X\{n,m\}$	Не менее n и не более m заглавных букв X. Если опустить m, то оператор означает «как минимум n заглавных букв X».
$(abc def)^+$	Не менее одной последовательности abc, def, abc и def будут считаться совпадением.

POSIX класс	Подобно	Означает
<code>[:upper:]</code>	<code>[A-Z]</code>	символы верхнего регистра
<code>[:lower:]</code>	<code>[a-z]</code>	символы нижнего регистра
<code>[:alpha:]</code>	<code>[A-Za-z]</code>	символы верхнего и нижнего регистра
<code>[:alnum:]</code>	<code>[A-Za-z0-9]</code>	цифры, символы верхнего и нижнего регистра
<code>[:digit:]</code>	<code>[0-9]</code>	Цифры
<code>[:xdigit:]</code>	<code>[0-9A-Fa-f]</code>	шестнадцатеричные цифры
<code>[:punct:]</code>	<code>[.,!?:....]</code>	знаки пунктуации
<code>[:blank:]</code>	<code>[\t]</code>	пробел и TAB
<code>[:space:]</code>	<code>[\t\n\r\f\v]</code>	символы пропуска
<code>[:cntrl:]</code>		символы управления
<code>[:graph:]</code>	<code>^[^ \t\n\r\f\v]</code>	символы печати
<code>[:print:]</code>	<code>^[^ \t\n\r\f\v]</code>	символы печати и символы пропуск

«Жадные» выражения

- «Ленивые» выражения
- «Жадные» выражения
- «Ревнивые»(сверхжадные) выражения

Жадный	Ленивый	Ревнивый
*	*?	*+
?	??	?+
+	+?	++
{n,}	{n,}?	{n,}+

Реализации

- NFA (Nondeterministic Finite State Machine; Недетерминированные Конечные Автоматы)
- DFA (Deterministic Finite-state Automaton; Детерминированные Конечные Автоматы)

GREP

- утилита командной строки («search globally for lines matching the regular expression, and print them»);
- находит на вводе строки, отвечающие заданному регулярному выражению, и выводит их, если вывод не отменён специальным ключом;
- изначально была создана для операционной системы UNIX;

Модификации GREP

- egrep (с обработкой расширенных регулярных выражений):
 - `grep -E`
- fgrep (трактующая символы `$*[]^|()\` буквально):
 - `grep -F`
- rgrep (с включённым рекурсивным поиском):
 - `grep -r`

Синтаксис GREP

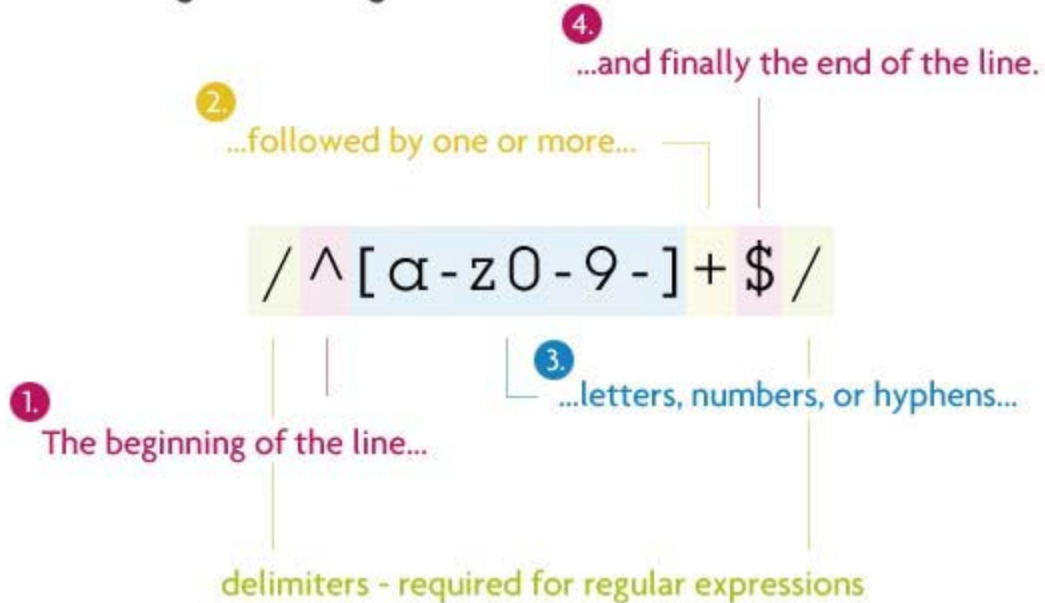
- `grep 'регулярное выражение' 'файл'`
 - `grep -E '^(bat|Bat|cat|Cat)' heroes.txt`
 - `grep -i -E '^(bat|cat)' heroes.txt`
- `cat 'файл' | grep 'регулярное выражение'`
 - `cat heroes.txt | grep -E '^[bcBC]at'`

Примеры для GREP

- Маркирование
`grep --color t[a-z] midsummer`
- Отображение совпадений
`grep -o '[A-Za-z]+' midsummer`
- Дизъюнкция или множества
`grep -E '^[bcBC]at' heroes.txt`
`grep -E '(bat|Bat|cat|Cat)' heroes.txt`

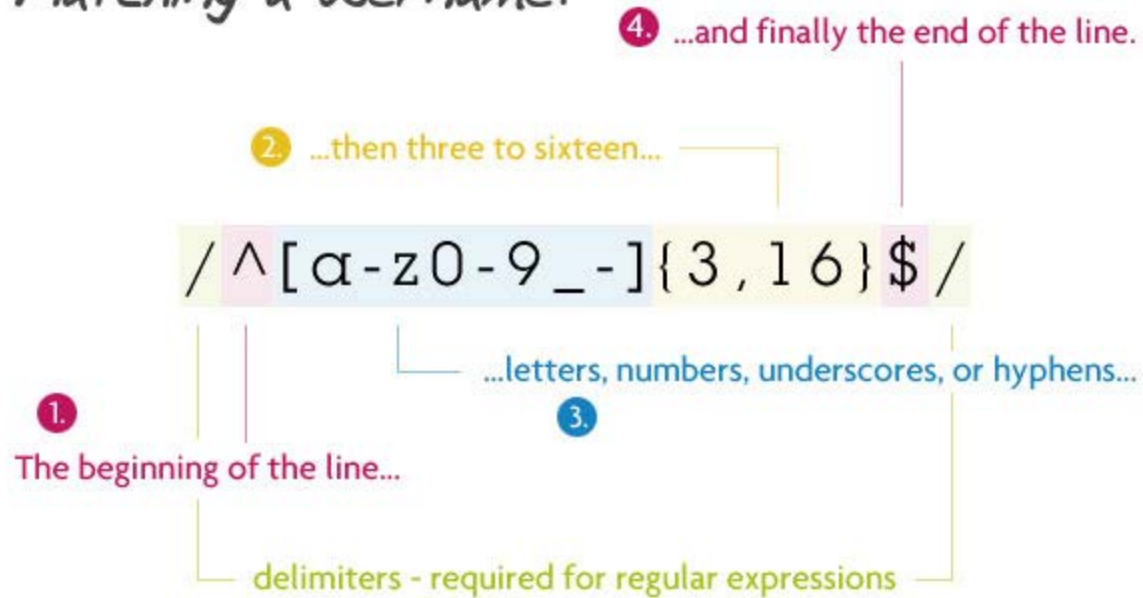
Пример 1. Слово с дефисами

Matching a "slug":



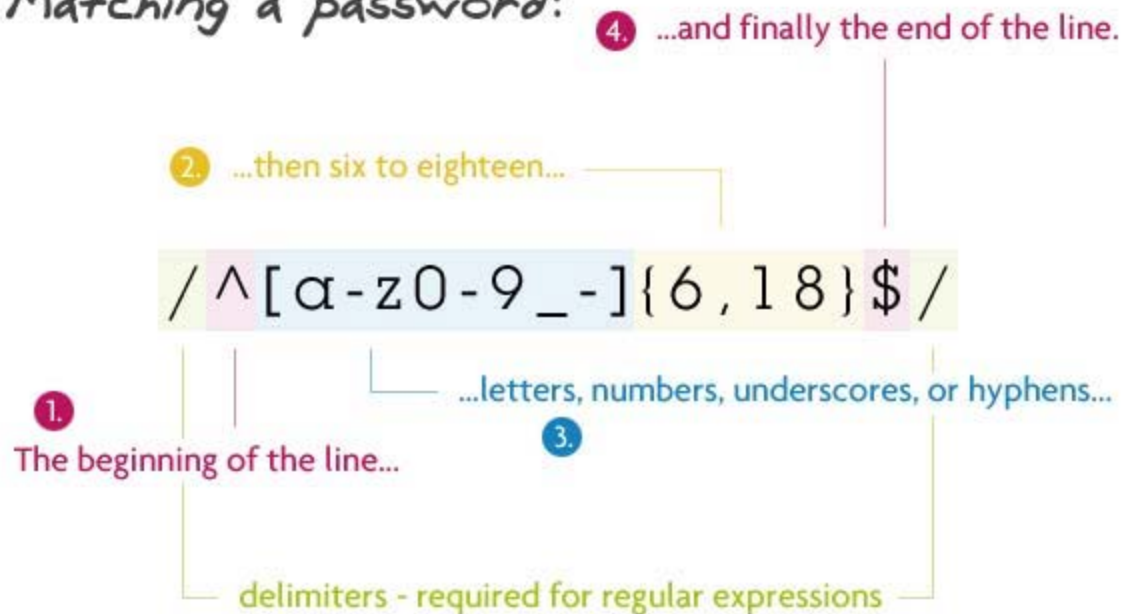
Пример 2. Username

Matching a username:



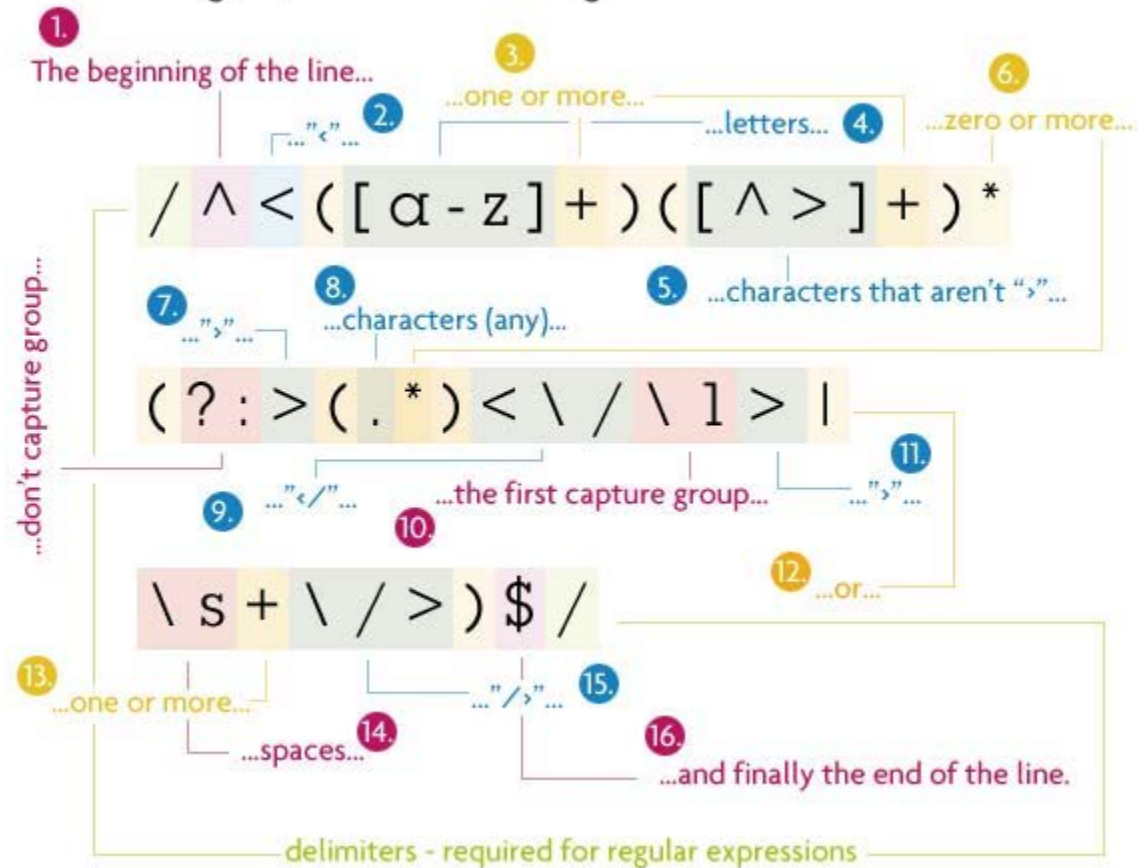
Пример 3. Пароль

Matching a password:



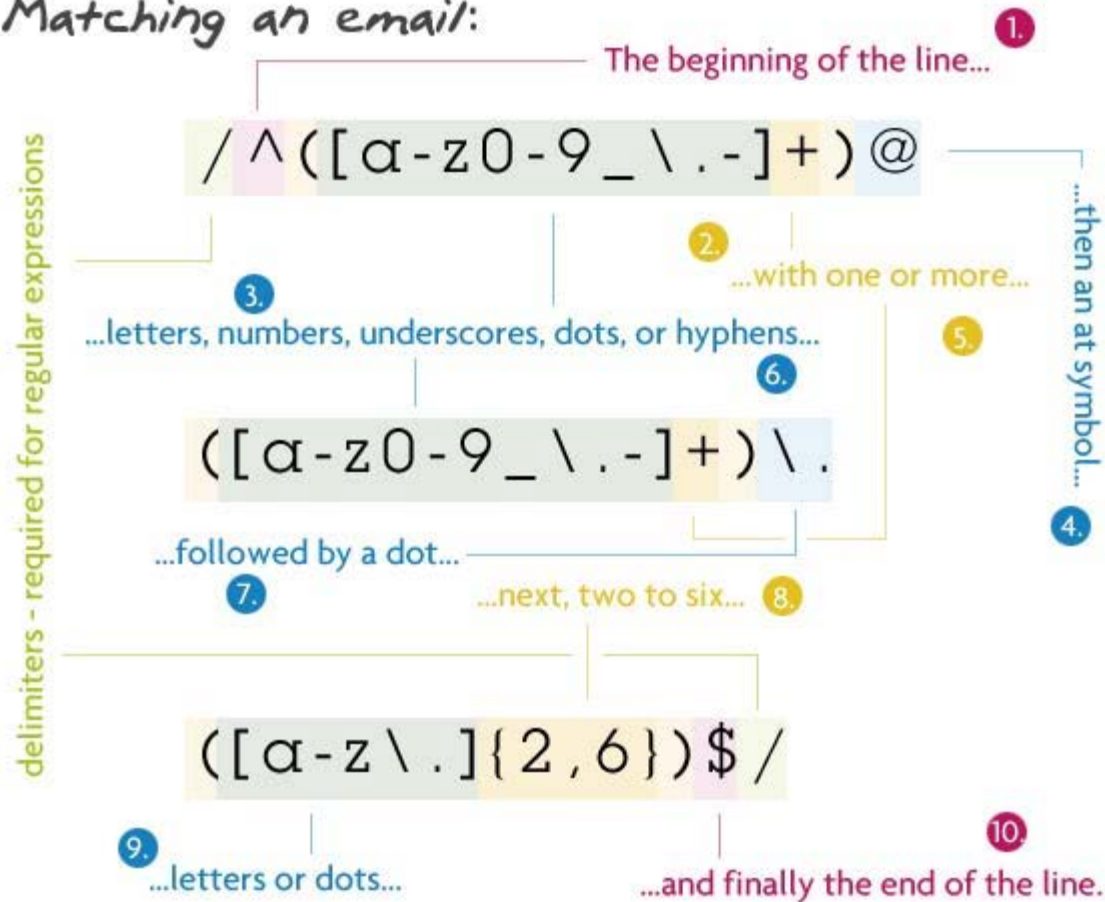
Пример 4. XML tag

Matching an HTML tag:



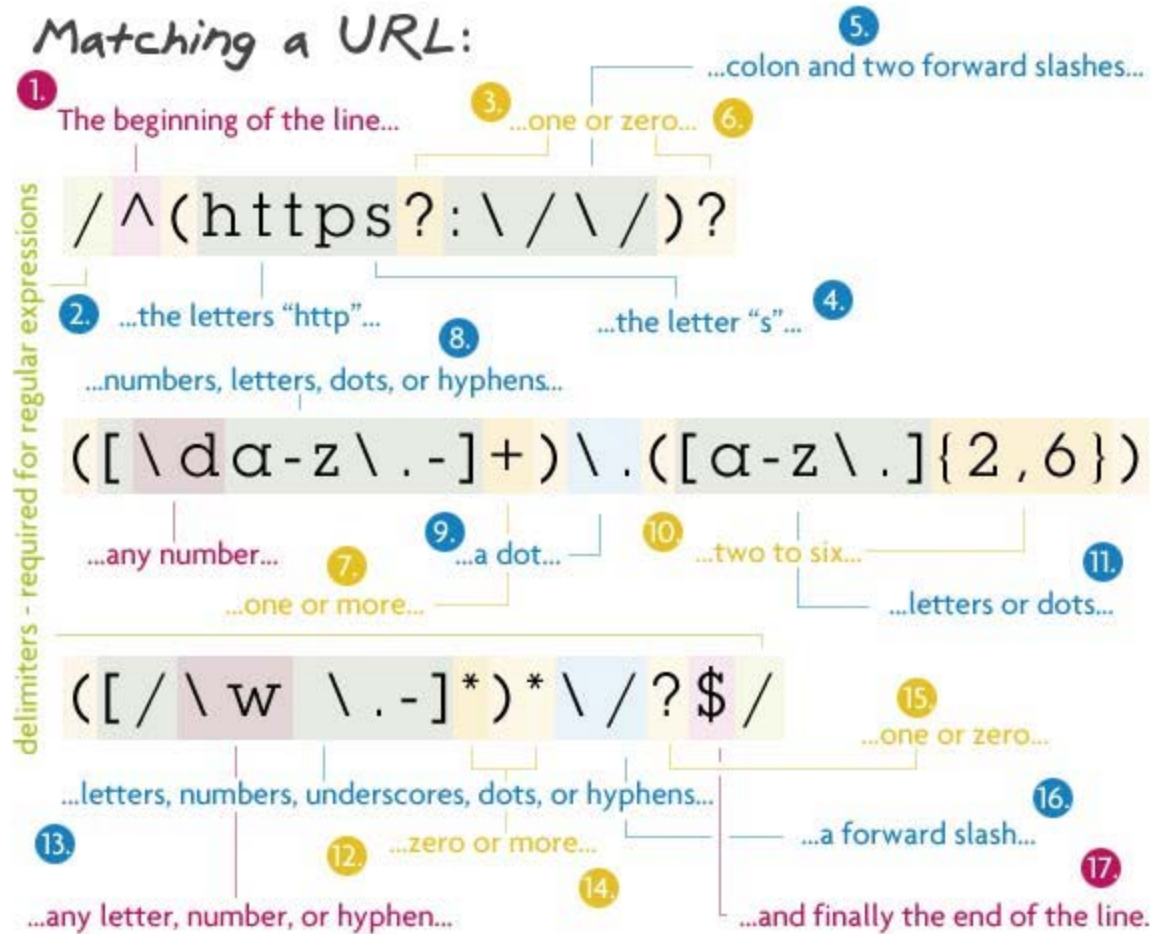
Пример 5. Email

Matching an email:



Pattern 6: URL

Matching a URL:



Пример 6 IP адрес

Matching an IP address:

The beginning of the line...

/ ^ (? : (? : 25 [0 - 5] |

...one or zero...

...don't capture group...

...any number...

2 [0 - 4] [0 - 9] | [0 1] ? [0 - 9] [0 - 9] ?)

...any number between zero and four...

...the number zero or one...

... "or" ...

\ .) { 3 } (? : 25 [0 - 5] | 2 [0 - 4] [0 - 9] |

...exactly three...

... "25" ...

... "2" ...

[0 1] ? [0 - 9] [0 - 9] ?) \$ /

...and finally the end of the line.

delimiters - required for regular expressions