

1. Регулярное выражение для поиска почты с русским доменом

```
\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.(рф|рус|москва|дети|онлайн|сайт|ком|орг)\b
```

- \b — граница слова
- [A-Za-z0-9._%+-]+ — имя пользователя
- @ — символ “собака”
- [A-Za-z0-9.-]+ — доменное имя
- \.(рф|рус|москва|дети|онлайн|сайт|ком|орг) — русские (и дополнительно латинские) домены

Пример использования в grep:

```
grep -E -o '\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.(рф|рус|москва|дети|онлайн|сайт|ком|орг)\b' файл.txt
```

2. Вытащить все IP-адреса пользователей

```
grep -Eo '\b([0-9]{1,3}\.){3}[0-9]{1,3}\b' файл.txt
```

Или если IP встречаются только в логах и идут после слов вроде “user” или “from”:

```
grep -Po '(?<=user ).*?\d{1,3}\.){3}\d{1,3}' файл.log
```

3. Найти файлы в директории, содержащие genome или GCA в названии

```
find /path/to/dir -type f \( -iname '*genome*' -o -iname '*GCA*' \)
```

4. Анализ GFF-файла

Формат строки GFF:

```
chr1 source feature start end score strand frame attributes
```

а) Извлечь гены:

```
awk -F '\t' '$3 == "gene"' файл.gff > genes.gff
```

б) Фильтр по признакам:

```
awk -F '\t' '
$3 == "gene" && $7 == "-" && match($9, /Name=([^;]+)/, a) && length(a[1]) > 3 &&
a[1] !~ /^[AZH]/ {
    print $1, a[1], $4, $5, $7
}
```

```
$3 == "gene" && $7 == "+" && match($9, /Name=([^;]+)/, a) && a[1] ~ /^[AZH]/ {
    print $1, a[1], $4, $5, $7
}
' OFS='\t' файл.gff > filtered_genes.tsv
```

с) Найти транскрипты по gene_id:

Сначала получить ID гена:

```
awk -F '\t' '
$3 == "mRNA" && match($9, /Parent=([^;]+)/, a) {
    print a[1], $9
}
' файл.gff > transcripts_by_gene.tsv
```

д) Объединить данные:

```
awk -F '\t' '
$3 == "gene" {
    chr = $1
    len = $5 - $4 + 1
    strand = $7
    chr_gene_count[chr]++
    chr_gene_len[chr] += len
    strand_count[strand]++
}
END {
    printf "Chromosome\tGene_Count\tAvg_Length\n"
    for (chr in chr_gene_count) {
        avg = chr_gene_len[chr] / chr_gene_count[chr]
        printf "%s\t%d\t%.2f\n", chr, chr_gene_count[chr], avg
    }

    printf "\nStrand\tGene_Count\n"
    for (s in strand_count) {
        printf "%s\t%d\n", s, strand_count[s]
    }
}
' файл.gff
```