

Драги читатели,

Понастоящем количеството информация в света расте експоненциално. Проблемът е, че се претоварваме от това, което идва до нас и предизвикателството е да се отдели наистина важното, възприемайки точната доза знание в точния момент, като се разграничат верните и важни факти от неточните. Очаква се през 2020 г. обемът информация в света да достигне 45 ZB (45 дзетабайта), които се равняват на 45 трилиона гигабайта (GB). Ако тази информация се складира в *i*-падове, всеки от които е с 128 гигабайта памет, то бихме получили купчина *i*-падове с височина, равна на 6 пъти разстоянието от Земята до Луната.

Да припомним мерните единици:

- 1 KB (килобайт) = 1000 байта;
- 1 MB (мегабайт) = 1000² байта;
- 1 GB (гигабайт) = 1000³ байта;
- 1 TB (терабайт) = 1000⁴ байта;
- 1 PB (петабайт) = 1000⁵ байта;
- 1 EB (ексабайт) = 1000⁶ байта;
- 1 ZB (дзетабайт) = 1000⁷ байта;
- 1 YB (йотабайт) = 1000⁸ байта.

(Байтът е единица за цифрова информация, която се състои от 8 бита, т.е. от 8 състояния на нули или единици.)

Част от информацията е във формат, който може да се възприема само от хората – статии, списания, документи, диаграми, снимки. В този си вид информацията е известна като неструктурирани данни, защото не е подредена в редове и стълбове, за да може да се обработва с помощта на софтуер. С такава информация може да се процедира толкова бързо, колкото бързо хората могат да четат, т.е. бавно. Но на помощ идва машинното учене, свързано със способността на компютрите да се самообучават без да са изрично програмирани за това. Машинното самообучение е дял от компютърните науки, който еволюира от разпознаване на образи и обучителна теория към изкуствен интелект. През 1959 г. Артър Самюел определя машинното самообучение като поле на проучване, което дава на компютрите възможност да учат. Към днешна дата машинното самообучение е отишло значително напред. То изследва проучването и изграждането на алгоритми за правене на прогнози чрез данни, от които да се учи. Машинното самообучение отключва способността за обработка на неструктурирани данни и стартира извличането на прозрения от тях. Използването му спестява десетки човеко-години работа и резултатите идват в рамките на няколко месеца.

Вече са създадени и продължават да се създават специални платформи, чрез които човек получава информация във връзка с направени заявки, но получава също и информация, която евентуално би била от негов интерес независимо, че може да е свързана с напълно различна област. Платформата предлага, например няколко статии или даже само няколко параграфа, или снимки, които току-що са били публикувани и които биха имали отношение към интересите на заявителя. Технологиата не е проблем, проблем са данните. Стартира се с данни, те се обработват във връзка с поставена задача и се подбират нови данни, необходими за решението. Затова може да се каже, че днес данните се превръщат в реална валута на комерсиалния свят.

Д-р М. Плюс