

ЖУРНАЛИСТИКА С БАЗИ ДАННИ

Живеем във време, в което сме затрупани от информация. Навсякъде, по всяко време и където и да се намираме, можем да намерим информацията, която ни интересува. Но наистина рядко си задаваме въпросът как да използваме всичката тази информация.

Заобиколени сме от хиляди камери, които ни записват - държавни, неправителствени и търговски организации събират и съхраняват данните ни в безброй бази данни. Исторически и статистически данни, лични данни, транзакционни и поведенчески данни са складирани в огромни масиви и трудни за асимилиране от повечето хора.

Тук идва мястото на журналистиката с бази данни, която трябва да превърне тези данни в полезна информация. Какво представлява журналистиката с бази данни? Най-просто казано този нов вид журналистиката се прави с данни. Но това е твърде повърхностно и твърде просто обяснение.

Все още битува понятието, че „данни“ са всяко множество от числа, събрани в електронна таблица. И наистина, преди 20 години това беше единственият вид данни, с които работеха журналистите. Но ние живеем в дигитален свят, свят, в който почти всичко може да бъде - и почти всичко е - описано с цифри – сивито ни, конфиденциалните документи, които подписваме, снимки, видео и аудио. Всичко това е описано с едни и същи две числа: нули и единици. Убийства, болести, политически гласове, корупция и лъжи: отново нули и единици.

Какво прави журналистиката с бази данни различна от останалата част от журналистиката? Това е уникалната комбинация от „нюх

за новини“ и способност журналистът да разкаже и онагледи завладяваща история, с мащаба и обхвата на огромната цифрова информация, която вече е на разположение.

Тези възможности могат да дойдат по различни пътища до журналиста: чрез използване на програми за автоматизиране на процеса на събиране и комбиниране на информация от местните власти, полицията и други източници, както е направил Адриан Холовати с ChicagoCrime и след това EveryBlock. Или чрез използването на софтуер, за да се намерят взаимовръзки между стотици хиляди документи, така както The Telegraph е направил с разходите на депутатите.

Информационната журналистика може да помогне на журналист да разкаже сложна история чрез инфографика, както направи Ханс Рослинг с презентацията си за бедността в световен мащаб с Garminder и предизвика милиони коментари по света.¹

Данните могат да бъдат източник или може да са инструментът, с който се разказва историята, а могат да бъдат и двете. Както всеки източник, те трябва да бъдат третирани внимателно и още по-внимателно интерпретирани.

Още през 2010 Сър Тим Бърнърс-Лий заявява, че е видял бъдещето на журналистиката - и като се има предвид, че той е изобретателят на World Wide Web, има доста голяма вероятност наистина да е така.

Бърнърс-Лий смята, че бъдещето е в анализа на данни. Много данни. При първото публикуване на данни от британското правителство за разходите от министерства в размер над 25 000 лири, той бе запитан кой ще ги анализира, след като компютърджиите са си намерили други занимания. Какъв е

¹ https://datajournalismhandbook.org/1.0/en/introduction_0.html; Paul Bradshaw; How to be a Data Journalist?"; 22,04,2019

смисълът от това? Кой реално ще държи правителството - или някой друг - отговорен за тези разходите?

"Отговорността би трябвало да е на пресата. Журналистите трябва да могат да боравят с данни. Някога можеше да се добереш до материали, като разговаряш с хората по баровете - и все още може и да е възможно да го правиш по този начин, поне понякога. Но сега трябва и ровене из данни и снабдяване с инструменти за анализа им и подбирането на това, което е интересно." отговаря Бърнърс-Лий.

Скандалът с разходите на свой ред доведе до откупено от Daily Telegraph специално проучване с участието на екип от хора, изучаващи съдържанието на данните. Последва публикуване на сведенията за разходите на депутатите, където медиите привлякоха самите читатели чрез специално създадени инструменти да се опитат да анализират какво се е случвало.

"Журналистиката, анализираща данни, е бъдещето," настоява Бърнърс-Лий. На което неговият колега Найджъл Шадболт, който заедно с Бърнърс-Лий се бори обществените служби и местната администрация да отворят достъпа до данните си, добавя кратко: "Е, поне част от бъдещето".²

КАК СЕ РОДИ ЖУРНАЛИСТИКАТА С ДАННИ?

Онлайн редакторите на британския вестник Гардиън са тези, които задвижват нещата през 2010 г., като обработват хиляди секретни документи за войната в Афганистан, които са им изпратени от платформата за разкрития Уикилийкс. От над 90 000 записа със секретни документи, те успяват да разработят мултимедийно обогатени доклади и интерактивни графики. В допълнение вестникът създава и база данни в интернет, която

² <https://www.goethe.de/ins/bg/bg/kul/mag/20744960.html>; Нора С. Штамплфл; „Данните разказват истории“ 22,04,2019

потребителите могат да ползват за правене на свои разследвания. По-късно Гардиън определя този свой начин на действие като неизбежен: обичайните методи като текстуализация или табличните форми наистина не са били подходящи за установяване на информационното съдържание в такъв огромен и хаотичен масив от данни.³

НУЖДАЕ ЛИ СЕ ЖУРНАЛИСТИКАТА С ДАННИ ОТ ЖУРНАЛИСТИ?

Краткият отговор е ДА! Журналистите, които работят с бази данни, работят съгласно същите етични и професионални принципи като традиционните журналисти. Понеже журналистите, които работят с данни, разчитат на специфични източници на материал и на специфични инструменти, тяхната роля също е по-различна: ако по-рано журналистите възприемаха себе си като пазители на източниците си, то при журналистиката с данни от тях се изисква много по-голяма откритост. Причината е, че журналистиката с данни се основава на принципите на една свободна и отворена интернет култура. От особено значение е също и идеята, че, чрез споделяне на данните с потребителите, те ще изпратят обратно нова информация към редакцията на медията, и така ще се стигне до по-голямо изясняване и разбиране на въпросната тема. Освен това, журналистите трябва да имат добри технически умения, за да съумеят да се възползват от възможностите на интернет и да направят едно качествено разследване. Разследванията в бъдеще често ще бъдат правени и чрез „scraping“ (остъргване),

³ <http://rhetoric.bg/%D1%81%D1%82%D0%B5%D0%BB%D0%B0-%D0%B0%D0%BD%D0%B3%D0%BE%D0%B2%D0%B0-%D1%82%D1%80%D0%B0%D0%BD%D1%81%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8-%D0%BD%D0%B0-%D0%BC%D0%B5%D0%B4>; Стела Ангова; „Трансформации на медийната среда“ 22,04,2019

т.е. посещаване на уеб сайтове и събиране на съществуващите данни.

Всичко това все пак не означава, че журналистите се превръщат в „стъргачи на данни“, статистици и програмисти. Не означава също, че в бъдеще журналистиката няма да се нуждае от журналисти, тъй като програмите могат автоматично да правят заявки за бази данни, да ги обработват и след това да ги възпроизвеждат. Тъй като данните не говорят сами за себе си, и занапред ще има нужда от класическите журналистически рамки: едва когато източниците на данни бъдат изяснени, взаимовръзките бъдат посочени и резултатите анализирани, тогава масивите от данни могат да развият една информационна добавена стойност.

Журналистиката на данните може да бъде разглеждана и като подвид на разследващата журналистика. Една от водещите медии в областта на журналистиката, основана на данни, е британският вестник Гардиън, който вече споменах и който има специален отдел, посветен единствено на обработката на данни за целите на журналистиката.

Много специалисти обръщат внимание на данните – това е ниша, в която ще се развиват доста компании. През 2014 г. компанията Farsite абсолютно вярно прогнозира носителите на наградата Оскар след обработка на гигантски обем данни за повече от 40-годишен период. Това насочва вниманието към възможността да се проследят най-популярните тенденции в реално време и да бъдат изведени определени модели от тях.

За разлика от традиционната журналистика, базирана на индивидуални и често субективни източници на информация, журналистиката с данни е базирана на множество данни от различни източници. Тъй като данните са най-често исторически,

тяхната манипулация е по-малко възможна, освен ако по някаква причина няма непълнота на наличните данни.

Наличието на данни за потребителите може да помогне на журналистите да анализират целевата си аудитория и така те да приоритизират нуждите и интересите на зрителите и читателите си. Взаимовръзките между анализираните параметри в данните могат да създадат модели за прогнозиране на бъдещи събития или поведение.⁴

НО КАКВИ СА ПРЕДИЗВИКАТЕЛСТВОТА ПРЕД ЖУРНАЛИСТИКАТА С ДАННИ?

Наличието и “чистотата” на данни е основното предизвикателство в работата с масиви от информация. След тях и освен техническите познания или инструменти се нарежда и разбирането на конкретния предмет на анализа.

Много често в многобройните бази данни конкретни параметри, предмет на анализ, които се въвеждат ръчно или без определен стандарт, не са въведени въобще или са въведени по различни начини, което затруднява работата или я прави невъзможна.

“Изчистването” на събраните данни и привеждането им към единен, “униформен” стандарт е ключово, ако искаме да получим ясна и обективна картина на база на наличната информация. Тази дейност често отнема значително време и за журналистите е изключително важно да проверят (валидират) окончателния резултат, за да ползват пълна и обективна информация. Например: при невъведени или въведени грешно възрастови данни (ЕГН или рождена дата) на конкретни целеви

⁴ https://www.webcafe.bg/id_1498248676- Бъдещето на журналистиката? В анализа на данни!; Webcafe.bg , По Guardian; 25.04.2019

групи медийна аудитория, не може да се направи конкретен извод за демографския им профил и съответно за интересите им.

Друго основно предизвикателство е необходимостта от набавяне на осъвременени данни при изготвянето на прогнозни поведенчески модели на база на исторически данни.

Изключителната динамика на променливите интереси, навици и поведение изисква постоянно набавяне на нови данни, тъй като профилите на анализирани целеви групи се променят постоянно.

Приложимостта на данните също е от ключово значение при анализа им и респективно – при изводите. Например анализ на потребностите на гражданин на Германия трудно би могъл да бъде приложен например във Вануату.

Ако събирането и изчистването на данни би могло да бъде автоматизирано, поне до голяма степен, то анализите и изводите от наличните данни остават изцяло в ръцете на журналистите.

Източници:

https://datajournalismhandbook.org/1.0/en/introduction_0.html;

Paul Bradshaw; How to be a Data Journalist?"; 22,04,2019

https://datajournalismhandbook.org/1.0/en/introduction_0.html;

Paul Bradshaw; What Is Data Journalism?

<https://www.goethe.de/ins/bg/bg/kul/mag/20744960.html>; Нора С.

Штампфл; „Данните разказват истории“ 22,04,2019

<http://rhetoric.bg/%D1%81%D1%82%D0%B5%D0%BB%D0%B0-%D0%B0%D0%BD%D0%B3%D0%BE%D0%B2%D0%B0-%D1%82%D1%80%D0%B0%D0%BD%D1%81%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8-%D0%BD%D0%B0->

[%D0%BC%D0%B5%D0%B4](#); Стела Ангова; „Трансформации на медийната среда“ 22,04,2019

https://www.webcafe.bg/id_1498248676- Бъдещето на журналистиката? В анализа на данни!; Webcafe.bg , По Guardian; 25.04.2019