



საქართველოს მეცნიერებათა ეროვნული აკადემია

ქართული ენციკლოპედიის ი. აბაშიძის სახელობის მთავარი სამეცნიერო რედაქცია

ინფორმაციის თეორია

ინფორმაციის თეორია, გამოყენებითი მათემატიკის დარგი, რომელიც შეისწავლის ინფორმაციის შეფასების, კოდირების, გადაცემის, შენახვის, მიღების, დეკოდირებისა და დახარისხების მათემატიკური აღწერის მეთოდებსა და საშუალებებს. ი. თ. წარმოიშვა ინფორმაციის შეკუმშვის ზომებზე და ხმაურიან არხზე ინფორმაციის საიმედოდ გადაცემის სისწრაფის ზღვრების დასადგენად. ი. თ., ძირითადად, შეისწავლის ინფორმაციის რაოდენობრივ ზომას.

ი. თ-ის, როგორც მეცნიერების, დასაწყისად ითვლება ამერ. მეცნიერის კ. შენონის სტატია „კომუნიკაციის მათემატიკური თეორია“ (1948), რ-შიც შემოღებულია ი. თ-ის ძირითადი ცნებები, მ. შ. ინფორმაციის რაოდენობის ერთეული ბიტი (ინგლ. bit, binary digit). ი. თ. გამოიყენება ინფორმაციის შეკუმშვისთვის დანაკარგების გარეშე (მაგ., ZIP ფაილების შესაქმნელად), დანაკარგებით (მაგ., MP3 მუსიკის ან JPEG სურათების ფორმატები), ხმაურიანი არხებით ინფორმაციის გადაცემისას [მაგ., მობილური ტელეფონები, ინტერნეტის DSL ხაზები, კოსმოსური ხომალდებიდან ინფორმაციის (სურათების) დაბრუნება დედამიწაზე] და სხვ.

ი. თ-ის გამოყენებით პოულობენ კოდირების მეთოდებს. იგი საშუალებას იძლევა, მოცემული ინფორმაცია აისახოს მაქსიმალურად შეკუმშულ კოდში, იმის გათვალისწინებით, რომ გადაცემისას შეიძლება მისი ნაწილი დაიკარგოს ან შეიცვალოს. ამისთვის კოდში უნდა იყოს დამატებითი სიმბოლოები, რ-ებიც არასრულად და (ან) არასწორად გადაცემული ინფორმაციის ამოცნობის საშუალებას იძლევა. მაგ., ქართულ ენაში (და სხვა ენებშიც) კავშირები, რ-ებიც ხშირად გვხვდება, მოკლე სიტყვებით გამოითქმის, ხოლო შედარებით იშვიათი სიტყვები გრძელია, რაც საშუალებას იძლევა

ვიმეტყველოთ მოკლე წინადადებებით. მეორე მხრივ, ჩვეულებრივი ენა „ჭარბად ინფორმატიულია“, სიტყვებში ბევრი „ზედმეტი“ სიმბოლოა, რაც საშუალებას იძლევა შეცდომების შემთხვევაშიც კი ამოვიცნოთ საჭირო ინფორმაცია. შესაბამისად, ი. თ-ის ამოცანაა, ინფორმაციის იმ მინ. რაოდენობის დადგენა, რ-საც უნდა შეიცავდეს ტექსტი (სიტყვები, სიმბოლოები და ა.შ.), რათა მისი ამოცნობა შეცდომებით გადაცემის შემთხვევაშიც იყოს შესაძლებელი. დანაკარგების გარეშე შეკუმშვის მაგალითია მორზეს კოდი. უფრო ხშირად ხმარებული ასოები აქ წარმოდგენილია მოკლე კოდებით (მაგ., „ე“ – მხოლოდ წერტილით), ხოლო შედარებით იშვიათი ასოები – უფრო გრძელი კოდებით. ეს საშუალებას იძლევა სიტყვები მოკლე კოდებით გამოისახოს.

მოვლენის შესახებ ინფორმაციის რაოდენობა I გამოითვლება ჰარტლის ფორმულით: $I = \log_{1/2} p = -\log_2 p$ (1928; $I = \log(p)$ p მოვლენის მოხდენის ალბათობაა). თუ მომხდარ მოვლენათა ალბათობებია p_1, \dots, p_k , მაშინ $I = -(\log_2 p_1 + \dots + \log_2 p_k) = -\sum (p_i \log_2 p_i)$ (შენონის ფორმულა, 1948). თერმოდინამიკური ენტროპიისთვის ცნობილ ბოლცმანის ფორმულასთან ამ ფორმულის მსგავსების საფუძველზე ხშირად იხილავენ ენტროპიას ი. თ-ში, როგორც ინფორმაციის არქონის ზომას. ამ მიდგომის თანახმად, ინფორმაციის ნებისმიერი რაოდენობის მოსაპოვებლად აუცილებელია შესაბამისი რაოდენობის ენერჯის დახარჯვა, ანუ ინფორმაციას აქვს გარკვეული „ენერგეტიკული ფასი“.

ინფორმაციის რაოდენობის ერთეული ბიტი (1 ბიტ) შეესაბამება ინფორმაციას ისეთი მოვლენის მოხდენის შესახებ, რ-ის ალბათობაა $1/2$; $8 = 2^3$ ბიტ შეადგენს 1 ბაიტს (ბტ). ი. თ-ში გამოიყენება უფრო დიდი ერთეულებიც (იხ. ცხრილი): $1024 = 2^{10}$ ბტ = 1 კბტ (KB), 2^{20} ბტ = 1 მეგბტ (MB).

თუ მოვლენის ალბათობაა 1 ან 0, მაშინ შესაბამისი ინფორმაციის რაოდენობაა 0. ქართულ 33-ასოიან ანბანში თითოეული ასოს ინფორმაცია საშუალოდ (თუ ჩავთვლით, რომ ყველა ასო ერთნაირი სიხშირით იხმარება) შეადგენს დაახლოებით 5 ბიტ ($33 \gg 2^5 = 32$). UNICODE მანქანური კოდების სისტემა შეიცავს 2^{16} სიმბოლოს; თითოეული სიმბოლოს ინფორმაცია შეადგენს 16 ბიტ = 2 ბტ. ტექსტში არსებული ინფორმაციის სრული რაოდენობა ტოლია სიმბოლოთა სრული რაოდენობის ნამრავლისა ერთი სიმბოლოს „ფასზე“ – სიმბოლოში ინფორმაციის რაოდენობაზე.

ი. თ-ის თანახმად, ინფორმაციის გაცვლა მიმდინარეობს სამ საფეხურად: კოდირება (ციფრულ ფორმატში გადაყვანა); გადაცემა (ამა თუ იმ ფიზიკური ბუნების გადამტანით); დეკოდირება (ციფრული ფორმატიდან თავდაპირველი, მაგ., გამოსახულების ან ხმოვანი ფორმატის აღდგენა).

„ხმაურიანი“ გადამცემი არხის ინფორმაციული გამტარუნარიანობა (დროის ერთეულში გადაცემული ინფორმაციის რაოდენობა) განისაზღვრება შენონის თეორემით, ფორმულით $(P_{\text{სიგნ}}/P_{\text{ხმაურ}} + 1)$, სადაც f_m გადასაცემი სიხშირეების ზოლის სიგანეა, ხოლო $P_{\text{სიგნ}}$ და $P_{\text{ხმაურ}}$ - არხში სიგნალისა და ხმაურის სიმძლავრეები. კოდირებისა და დეკოდირების საფეხურზე ანათვალთა მინიმალურ აუცილებელ რაოდენობას განსაზღვრავს ნაიკვისტ-კოტელნიკოვ-შენონის თეორემა: ნებისმიერი ფუნქცია, რ-იც შეიცავს სიხშირეებს 0-იდან f_c -მდე, შეიძლება ნებისმიერი სიზუსტით გადავცეთ რიცხვთა მიმდევრობით, რ-თა გადაცემის შუალედია $(2f_c)^{-1}$ წმ.

ი. თ-ის განვითარებაში დიდი წვლილი შეიტანეს ქართველმა მეცნიერებმა: ი. ფრანგიშვილმა, თ. ვაშაყმაძემ (თსუ), ჰ. მელაძემ (წმ. ანდრია პირველწოდებულის სახ. ქართული უნ-ტი), ა. ფრანგიშვილმა (სტუ), გ. ხარატიშვილმა და სხვ.
