

# Fuzzy Logic în automatizarea procesului de analiza a datelor din GoogleAnalytics

Sergiu CHILAT, Victor BEȘLIU  
 Technical University of Moldova  
 chilatsergiu@gmail.com, vbeshliu@yahoo.com

**Abstract** — În acest articol este prezentată o metodă de automatizare a procesului de analiză a datelor din Google Analytics cu utilizarea logicii fuzzy. În urma implementării metodei, utilizatorului i se va prezenta cel mai util conținut în dependență de vârsta pe care o are. Deoarece, în Google Analytics, nu este pus la dispoziție nici un instrument de analiză automată, implementarea acestui model ar aduce o serie de avantaje, unul din cele mai importante fiind economia de timp

**Cuvinte cheie** — Fuzzy logic, web analytics, automatization, self tuning, fuzzyfication.

## I. INTRODUCERE

Colectarea, măsurarea și analiza datelor despre utilizatorii aplicațiilor web a devenit o realitate. Scopurile și obiectivele sunt diferite, printre care și cel de maximizare a profitului prin prezentarea produselor și conținuturilor conform intereselor utilizatorului. Două din cele mai populare motoare de căutare Google and Yandex, pun la dispoziție instrumente specializate: Google Analytics și Yandex.Metrica. acestea permit acumularea datelor despre utilizatorii sitului, precum ar fi, locația geografică, vârstă, preferințe și interese, comportament etc [1]. Aceste date pot fi utilizate pentru a optimiza situl, însă la moment, nu există metode sau instrumente care ar putea să automatizeze procesul de luare a deciziilor, astfel, această atribuție îi revine administratorului SEO (Search Engine Optimization). Sunt frecvente situațiile în care administratorul SEO execută aceleași operații de rutină în mod repetat, ceea ce conduce la descreșterea randamentului și adoptarea unor decizii de proastă calitate. În acest articol va fi cercetată posibilitatea implementării logicii fuzzy la automatizarea procesului de luare a deciziilor, cu scopul de a economisi timp și a elibera administratorul de la activități de rutină. Unul din compartimentele Google Analytics este vârsta utilizatorilor, în baza căruia se va efectua cercetarea și va fi propusă o metodă de analiză automată a datelor și luarea deciziei optime, utilizând funcția de maximizare a utilității în procesul de defuzzyficare.

## II. PROCESUL DE FUZZIFICARE

Google Analytics grupează datele obținute în 6 grupe (fig. 1), pe care le vom nota cu GV1-GV6 (variabile de intrare):

- GV<sub>1</sub> - de la 18 până la 24 ani;
- GV<sub>2</sub> - de la 25 până la 34 ani;
- GV<sub>3</sub> - de la 35 până la 44 ani;
- GV<sub>4</sub> - de la 45 până la 54 ani;
- GV<sub>5</sub> - de la 55 la 64 ani;
- GV<sub>6</sub> - peste 65 ani.

Însă, informația prezentată pe site, este împărțită în 3 grupe, deoarece acest lucru se face destul de simplu [2]:

- GV<sub>T</sub> - pentru persoane tinere;
- GV<sub>M</sub> - pentru persoane de vârstă medie;
- GV<sub>B</sub> - pentru persoane vârstnice.

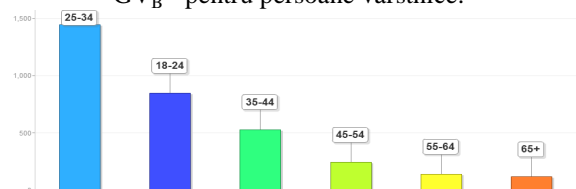


Fig. 1. Gruparea datelor după vârsta utilizatorilor în Google Analytics.

Deoarece nu există o limită de trecere clară între aceste categorii de vârstă, împărțirea în 6 grupe (GV1-GV6) pe care o pune la dispoziție Google Analytics, nu poate fi utilizată direct, este nevoie de elaborarea unui algoritm care ar determina din care grup din cele 3 (GVT, GVM, GVB) face parte utilizatorul.

Conform separării stricte pe care o face Google Analytics, persoana cu vârsta de 35 ani aparține grupei persoanelor de vârstă medie - GV<sub>M</sub>. O așa separare strictă, duce la micșorarea eficienței în expunerea informațiilor în pagină, deoarece în realitate, trecerea de la o grupă de vârstă la alta nu are loc brusc, ci este un proces continuu.

Utilizarea logicii Fuzzy [3] pentru prezentarea unui anumit conținut în dependență de grupa din care face parte utilizatorul elimină separarea strictă, iar în perspectivă poate duce la o expunere mai eficientă a informației.

În acest scop, se va defini variabila lingvistică vârstă cu valorile posibile: tînăr, mediu, bătrîn.

În tabelul 1 sunt prezentate punctele critice și valorile funcțiilor de apartenență în aceste puncte.

TABELUL 1. REPREZENTAREA FUZZY A DISTRIBUȚIEI VÂRSTĂ

	18	24	34	44	54	64	65+
Tînăr	1	0,75	0,25	0	0	0	0
Mediu	0	0,25	1	1	0,25	0	0
Bătrîn	0	0	0	0,25	0,75	1	1

În baza datelor din tabelul 1 se va construi diagrama fuzzy a distribuției variabilei vârstă (fig. 2).

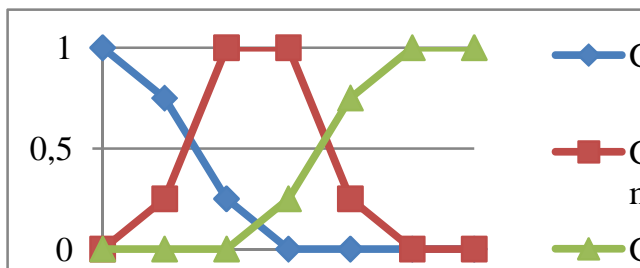


Fig. 2. Diagrama Fuzzy a distribuției variabilei vîrsta.

### III. DEFINIREA BAZEI DE REGULI

După cum se poate observa din fig. 2, un utilizator este "foarte" tînăr dacă are vîrsta de 18 ani (valoarea maximală 1), "mai puțin" tînăr la vîrsta de 24 ani (valoarea 0,8), "puțin" tînăr (valoarea 0,2) la vîrsta de 34 ani, iar la vîrsta de 44 ani nu mai este tînăr (valoarea 0).

La fel și pentru persoanele bătrîne, o persoană nu este bătrînă pînă la vîrsta de 24 ani (valoarea 0), "puțin" bătrînă la vîrsta de 34 ani (valoarea 0,2), "aproape" bătrînă (valoarea 0,8) la vîrsta de 54 ani, iar la vîrsta de 65 ani este bătrînă (valoarea maximală 1).

Pentru persoanele de vîrstă medie se observă o mică creștere (0,2) a funcției de apartenență de la 18 la 24 ani, și o creștere semnificativă în intervalul 24-34 ani (pînă la 0,8), iar în intervalul 34-44 este atinsă valoarea maximală (1). După 45 ani, valoarea funcției de apartenență descreește, ajungînd la 0 pentru vîrsta de 64 ani.

Baza setului de reguli [3, 4], care transformă datele de intrare în date de ieșire va fi de forma:

*IF age = young THEN content = for young*

*IF age = medium THEN content = for medium*

*IF age = old THEN content = for old*

Controlerul fuzzy propus este prezentat în figura 3. Intrarea sistemului (vîrsta) este supusă transformărilor dictate de setul de reguli, este stabilită categoria de vîrstă din care face parte utilizatorul, iar la ieșire va fi prezentat conținutul considerat cel mai util pentru utilizatorul dat.

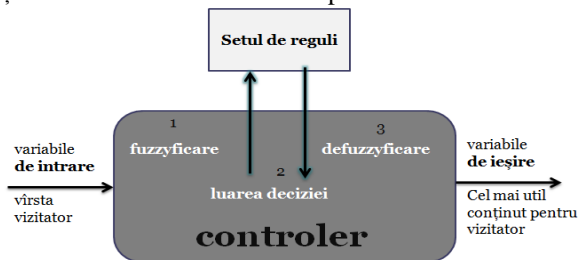


Fig.3. Controlerul fuzzy.

### IV. PROCESUL DE DEFUZZIFICARE

La această etapă variabila lingvistică content va indica sistemului din care grup să fie prezentat conținutul.

Dacă un conținut de pe site este destinat persoanelor de vîrstă medie, acesta va fi foarte util (va avea utilitatea maximală 1) pentru utilizatorii cu vîrsta din intervalul 34-44, pentru utilizatorii cu vîrste cuprinse între 24-34 și 44-54 va avea o utilitate medie (între 0,2 și 0,8) iar pentru utilizatorii cu vîrsta pînă la 24 ani și după 54 ani, acest conținut va avea o utilitate mică, deci nu prezintă interes.

Din acest motiv se va utiliza funcția de maximizare [5]  
 $optimalContent = MaxUtility(Content)$

deoarece scopul este de a prezenta conținutul cel mai relevant vîrstei utilizatorului.

Utilizînd datele, prezentate în fig. 1 și mulțimile fuzzy definite, se poate observa că numărul maxim de utilizatori au vîrsta cuprinsă între 25-34 ani și între 18-24 ani, deci, algoritmul va genera în mod automat două rezultate "Conținutul este util pentru persoanele de vîrstă tînără și medie" și "Nu este util pentru persoanele în vîrstă".

În cazul în care exista un conținut pentru care valoarea funcției de apartenență este 1, va fi prezentat acest conținut, în caz contrar va fi selectat un conținut cu valoarea maximă a funcției de apartenență.

Utilizarea acestui algoritm de automatizare este utilă în cazul siturilor cu un număr mare de pagini, astfel economisindu-se foarte mult timpul administratorului SEO.

### V. CONCLUZII

Metoda propusă permite selectarea automată a conținutului optim pentru a fi prezentat utilizatorului sitului. Avantajul principal este că, în cazul în care utilizatorul face parte din intervalul de vîrstă 25-34 (funcția de apartenență la mulțimea GM are valoarea maximă - 1), iar conținut pentru această grupă nu există, metoda propusă va selecta un conținut cît mai apropiat de interesele sale, adică din grupele învecinate.

În baza exemplului cercetat, se poate observa că datele, care inițial erau împărțite în mulțimi clasice, au fost regrupate în mulțimi fuzzy, după care a putut fi aplicată logica fuzzy pentru prelucrarea datelor și generarea soluțiilor. Crearea mulțimilor fuzzy este un prim pas în implementarea algoritmului bazat pe logică fuzzy, aceasta fiind una din cele mai importante etape, unde este nevoie de o generare corectă a acestor mulțimi.

Algoritmii fuzzy pot fi utilizați în rezolvarea problemelor nedeterministe, cu momente de incertitudine, acolo unde este nevoie de o "gîndire" asociativă și variabilă, asemenea gîndirii umane. Deoarece analiza datelor despre utilizatorii unui site este un proces continuu și neliniar, prelucrarea în timp real a datelor în mod automat oferă o serie de avantaje, cum ar fi: economie de timp, creșterea vitezei de procesare, excluderea greșelilor pe care le poate face administratorul SEO.

### REFERINȚE

- [1] Google Analytics Feature List, 2014. [http://www.google.ru/intl/en\\_uk/analytics/features/ind ex.html](http://www.google.ru/intl/en_uk/analytics/features/ind ex.html)
- [2] C. A. Smith. *Automated continuous process control* (2002), pp. 115-122.
- [3] L. A. Zadeh. *Fuzzy sets, Information and Control*, vol. 8 (1965), pp. 338-353.
- [4] L. A. Zadeh. *Computing With Words. Principal Concepts and Ideas* (2012), pp. 21-24.
- [5] C. A. Smith. *Automated continuous process control* (2002), pp. 115-122.