

(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2015 00607

(22) Data de depozit: 21/08/2015

(41) Data publicării cererii:
28/02/2017 BOPI nr. 2/2017

(71) Solicitant:
• IMINENT TECHNOLOGY S.R.L.,
STR. DIONISIE LUPU NR. 64-66, MEZANIN,
CAMERA 1, SECTOR 1, BUCUREȘTI, B,
RO

(72) Inventatori:
• CHIRU COSTIN- GABRIEL,
STR. EȘARFEI NR. 97A, BUCUREȘTI, B,
RO;
• ILAȘ CONSTANTIN,
STR. DUMITRU ZOSSIMA NR. 65,
BUCUREȘTI, B, RO

(54) METODĂ DE PREDICȚIE INDIVIDUALIZATĂ A
CATEGORIILOR PAGINILOR DE INTERNET CE URMEAZĂ
SĂ FIE VIZITATE DE UN UTILIZATOR

(57) Rezumat:

Invenția se referă la o metodă de predicție individualizată, bazată pe algoritmi de inteligență artificială, a categoriilor paginilor de internet care urmează să fie vizitate de un utilizator, metodă ce poate fi utilizată de browserele de internet și de furnizorii de servicii web, pentru a oferi sugestii de pagini, în vederea accelerării și simplificării navigației, cu beneficii pentru navigarea de pe calculatoarele personale, dar mai ales de pe dispozitive mobile. Metoda de predicție, conform invenției, constă în următoarele faze: o primă fază de preprocesare (3) în care, pornind de la o secvență curentă de navigare a utilizatorului, și de la istoricul său de navigare, se creează unul sau mai mulți vectori aciclici cu categoriile paginilor vizitate, în ordinea vizitării; o a doua fază de procesare (4), în care se prezice categoria următoarei pagini de internet care va fi accesată de utilizator, această predicție fiind efectuată combinând două tehnici de predicție diferite: una bazată pe folosirea unui procedeu de recomandare (12), iar cealaltă, o metodă statistică bazată pe compararea probabilităților diferitelor categorii de pagini de a fi accesate, și folosind trei factori diferiți pentru realizarea predicției: istoria navigării utilizatorului, momentul actual de timp și ritmul curent de navigare al utilizatorului; și o a treia fază, în care se adaugă informația preprocesată la istoricul de navigare a utilizatorului. Metoda de predicție furnizează la ieșire categoriile (O1) probabile la momentul actual de timp, categoriile (O2) probabile conform

ritmului de navigare curent, probabilitatea (O3) de continuare a secvenței, și categoria (O4) prezisă, prin combinarea categoriilor (O1, O2 și O3) anterioare.

Revendicări: 4
Figuri: 8

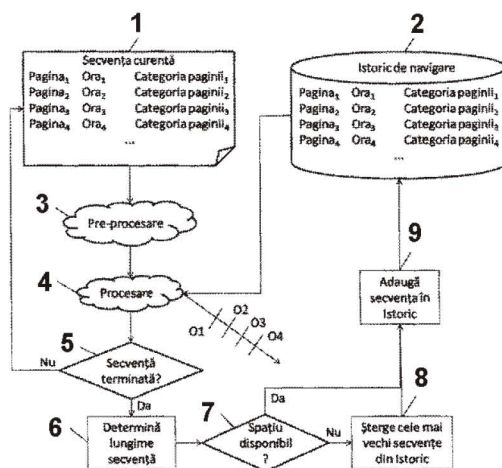


Fig. 1



Metodă de predicție individualizată a categoriilor paginilor de internet ce urmează să fie vizitate de un utilizator

Invenția se referă la o metodă de predicție, bazată pe algoritmi de inteligență artificială, a categoriilor probabile din care vor face parte paginile de internet ce urmează a fi vizitate de un utilizator, metodă care poate fi folosită de browserele de internet și de furnizorii de servicii web pentru oferirea de sugestii de pagini, în vederea accelerării și simplificării navigației. De asemenea, invenția poate fi folosită și pentru oferirea de servicii web complexe, bazate pe informații provenind din mai multe surse, obținute anticipat față de momentul când sunt cerute, precum și pentru creșterea securității, prin detectarea unui utilizator neautorizat. Prin categorie a unei pagini de internet se înțelege domeniul căruia îi aparține pagina, de exemplu: știri, filme, muzică, știință și tehnică, etc.

Sunt cunoscute diverse metode de predicție folosite în diferite aplicații. Unele dintre ele sunt specializate pentru dispozitive mobile, pentru sarcini ca predicția zonelor din documente ce trebuiesc pre-încărcate pe dispozitivele mobile, prezicerea cuvintelor ce se doresc a fi tastate pornind de la câteva caractere ambigue sau de la funcționalitățile telefoanelor, cum ar fi cele prezentate în documentele de brevet "Predicting and Retrieving Data for Preloading on Client Device", inventatori Chang J, Yuan S, Wang B, Tsai E, Jiang A (US2013304798A1, publicat 14.11.2013), "A Communication Terminal Having a Predictive Editor Application", inventator Haestrup Jan (EP1031914A2, publicat 30.08.2000), "Messaging System", inventatori Ezra D și Kaluaratchie D C (GB2470585A, publicat 01.12.2010).

Cele mai multe metode însă sunt fie specializate pentru dispozitive desktop, fie adaptate pentru ambele tipuri de dispozitive, pentru sarcini ca îmbunătățirea rezultatelor furnizate de motoarele de căutare, prezicerea următorului link ce urmează a fi vizitat de către utilizator, îmbunătățirea navigării pe internet prin pre-încărcarea de pagini web sau prin realizarea unui sistem de caching pe partea de server, prezicerea căilor de navigare alcătuite din mai mulți pași, prezicerea sit-urilor web ce se doresc a fi accesate folosind doar primele caractere tipărite într-un browser, evaluarea acțiunilor posibile ale utilizatorilor în cazul sistemelor IPTV, cum ar fi cele prezentate în documentele de brevet "Modelling, Detecting and Predicting User Behavior with Hidden Markov Models", inventator Heck Larry P. (US2009164395A1, publicat 25.06.2009), "System and Method for Auto-sensed Search Help", inventator Watson Eric B. (US2006259861A1, publicat 16.11.2006), "Predicting User Navigation Events Based on Chronological History Data", inventatori Burkard T, Hamon D și Jain A (US8655819B1, publicat 18.02.2014), "Accelerating User Interfaces by Predicting User Actions", inventatori Fredricksen E R., Buchheit P și Rennie J G. (WO2006085314A2, publicat 30.06.2004), "Access to Network Content", inventatori Burkard T și Jain A (US2014095966A1, publicat 03.04.2014), "Predicting User Navigation Events", inventatori Hamon D, Burkard T și Jain A (US8566696B1, publicat 22.10.2013), "Predicting User Navigation Events in an Internet Browser", inventator Hamon D (US8744988B1, publicat 03.06.2014), "A Novel Book-like Internet Browser for Electronic Information", inventatori Seet Chern H și Ho Seng B (WO0206917A2, publicat 24.01.2002), "Dynamic Prefetching Method and System for Metadata", inventatori Cetin G, Gumus U și Kupusoglu O (EP2194471A1, publicat 09.06.2010), "Web Site Connecting System Using Keyword and

Method Threerby", inventator Park Ho-nam (**WO2007027002A1**, publicat 08.03.2007), "Enabling Predictive Web Browsing", inventatori White C A. și Jones C D. (**US2012137201A1**, publicat 31.05.2012), "User Intention Modeling for Web Navigation", inventatori Zheng C, Xiaoming S și Liu W (**US2003212760A1**, publicat 13.11.2003), "Dynamic Page Generation Acceleration Using Component-Level Caching", inventator Anindya D (**US6622168B1**, publicat 16.11.2003), "Access to Network Content", inventatori Burkard T, Jain A și Bentzel C (**US2012324043A1**, publicat 20.12.2012), "Adaptive Prefetching for Computer Network and Web Browsing with a Graphic User Interface", inventatori Jiang Z și Kleinrock L (**US6385641B1**, publicat 07.05.2002), "Access to Network Content", inventator Hamon Dominic (**US8954524B1**, publicat 10.02.2015), "Web Brower Graphics Management ", inventatori Eves D și Timms A (**WO0019336A1**, publicat 06.04.2000), "Navigating Organizational Structures", inventatori Forman G H și Suermond H J (**US2003139901A1**, publicat 24.07.2003), "Accelerating User Interfaces by Predicting User Actions", inventatori Fredricksen E R., Buchheit P și Rennie J G. (**WO2006012144A2**, publicat 02.02.2006).

Aceste metode au multe dezavantaje. Un dezavantaj constă în aceea că, în toate cazurile, pentru predicție sunt folosite metode simpliste ce folosesc informații puține (bigrame sau cel mult n-grame de vizite de pagini) și presupuneri simpliste (Naive Bayse) pe baza cărora se pot deriva informații limitate (în general numai predicția următoarei pagini sau predicția unei secvențe de pagini, dar cu acuratețe din ce în ce mai scăzută).

Majoritatea metodelor de mai sus (cu excepția US2003212760A1) consideră drept istoric de navigare numai pagina curentă, fără a se raporta la istoria anterioară de navigare, iar aceasta istorie foarte scurtă nu permite obținerea unei precizii bune. Un alt dezavantaj constă în faptul că aceste metode folosesc informația referitoare la pagina curentă pentru a prezice numai următoarea pagină ce va fi accesată și nu o anumită cale de navigare. De cele mai multe ori, predicția se bazează pe parsarea paginii curente pentru a extrage legăturile existente către alte pagini web iar predicția consideră ca posibilități de continuare a navigării numai aceste legături, ceea ce necesită mai mult timp de procesare pentru a extrage legăturile din pagină și de aici un alt dezavantaj. Excepție fac US2003212760A1, US6622168B1. Unele dintre metodele menționate folosesc informații sensibile pentru clienți, ceea ce ridică probleme de încredere din partea acestora, reprezentând un alt punct slab.

În plus, metoda folosită în US2003212760A1, care apare ca excepție în lista de mai sus, are dezavantajul că are nevoie pentru a funcționa de o listă lungă de informații pentru care se păstrează istoria (url, moment de acces al acestuia, durata de timp petrecută pe fiecare pagină, modul în care s-a ajuns la un anumit url: în urma unui click pe un link, prin căutare folosind un motor sau direct scriind url-ul în browser, ce motor de căutare a fost folosit, cuvintele căutate, rezultatele obținute, pe care din rezultate întoarse de motorul de căutare s-a dat click) ceea ce înseamnă că, la implementarea pe un dispozitiv mobil, fie toată această informație amplă e salvată pe dispozitivul mobil (ocupând memoria) și atunci prelucrarea se face local (cu impact asupra duratei de viață a bateriei și a performanței dispozitivului), fie, dacă informația e transmisă pe un server, se va realiza mult trafic de date, cu impact negativ asupra duratei de viață a bateriei și a performanței dispozitivului, inclusiv a vitezei de navigare.

21-08-2015

Un alt dezavantaj major este faptul că, de multe ori, predicția pentru un utilizator se face pe baza acțiunilor anterioare ale altor utilizatori, deci nu este individualizat pentru fiecare utilizator în parte, ci ține cont de acțiunile majorității.

Problema tehnică pe care o rezolvă invenția este accelerarea și simplificarea navigației, prin oferirea, la momentul potrivit de timp, de sugestii pertinente, ușor de accesat, care este benefică indiferent de utilizator și de platforma folosită, dar este foarte importantă în cazul folosirii dispozitivelor mobile (telefoane, tablete). În acest caz, accesarea unor pagini conținând informația dorită la un moment dat se poate face prin folosirea motoarelor de căutare, dar aceasta presupune introducerea manuală a mai multor termeni, și eventual repetarea operației de câteva ori, până la găsirea a ceea ce se caută. Prin metoda propusă, aceasta etapă este eliminată, utilizatorul putând primi direct sugestii de pagini aparținând categoriei prezise. În plus, prin felul în care este concepută, metoda poate fi cu ușurință implementată, într-o manieră optimizată, atât pe dispozitive mobile (cu minimizarea puterii consumate și un volum redus de trafic de date suplimentar), cât și pe calculatoare personale.

Metoda de predicție, conform invenției, rezolvă problema tehnică enunțată și înlătură dezavantajele menționate, prin aceea că furnizează, cu probabilitate crescută și într-o manieră eficientă, intenția de navigare a unui utilizator, reprezentată prin categoriile paginilor de internet ce urmează să fie vizitate de către acesta, destinată utilizării pe calculatoare personale și dispozitive mobile (telefoane, tablete), în scopul accelerării și simplificării navigației, prin oferirea, la momentul potrivit de timp, de sugestii pertinente de pagini web, ușor de accesat de către utilizator. În acest fel, este posibilă preîncărcarea unei serii de pagini din această categorie și oferirea lor ca opțiuni de continuare a sesiunii de navigare, astfel permițând accelerarea și simplificarea navigației.

Metoda de predicție, conform invenției, constă în următoarele fazele:

- a. Pre-procesare, în care, pornind de la secvența curentă de navigare a utilizatorului și de la istoricul său de navigare se crează unul sau mai mulți vectori aciclici cu categoriile paginilor vizitate, în ordinea vizităților;
- b. Procesare, în care se prezice categoria următoarei categorii de pagini de internet ce va fi accesată de utilizator. Această predicție se face combinând două tehnici diferite de predicție: una bazată pe folosirea unui procedeu de recomandare, iar cealaltă, o metodă statistică bazată pe compararea probabilităților diferitelor categorii de pagini de a fi accesate; și folosind trei factori diferiți pentru realizarea predicției: istoria navigării utilizatorului, momentul actual de timp și ritmul curent de navigare al utilizatorului;
- c. Adăugarea informației pre-procesate în istoricul de navigare al utilizatorului.

Metoda de predicție prezentată, conform invenției, are următoarele avantaje:

- Are o precizie ridicată, deoarece:
 - o Este aplicată fiecărui utilizator în parte, fiind bazată pe prelucrarea informației de navigare specifică acestuia și nu prin medierea informațiilor provenite de la un grup mare de utilizatori;
 - o Combină două tehnici diferite de predicție: una bazată pe folosirea unui procedeu de recomandare, iar cealaltă, o metodă statistică bazată pe compararea probabilităților diferitelor categorii de pagini de a fi accesate;

- o Combină trei factori diferiți pentru realizarea predicției: istoria navigării, momentul actual de timp și ritmul curent de navigare.
- Folosește doar informații non-intruzive, puțin sensibile, adică momentul navigării, pagina de web accesată și categoria acesteia, și nu informații despre gen, vârstă, venit, pasiuni, locație, etc.
- Necesită ca informație de intrare doar momentul navigării, url-ul paginii accesate și categoria ei, deci permite cu ușurință implementarea optimă pe dispozitive mobile, în cazul cărora metoda e implementată pe server și primește aceste informații, fără impact important asupra traficului de date, duratei de viață a bateriei și performanțelor dispozitivului mobil.

Invenția este prezentată în legătură cu următoarele figuri, care reprezintă:

- fig. 1, principalele etape ale procedurii vizat de acest patent;
- fig. 2, diagrama de flux a procesării implicate de procedeu vizat de acest patent;
- fig. 3, diagrama de flux a determinării categoriilor probabile în funcție de diferite momente de timp;
- fig. 4, diagrama de flux a determinării categoriilor probabile la momentul curent de timp;
- fig. 5, diagrama de flux a determinării categoriilor probabile în funcție de ritmul de navigare;
- fig. 6, diagrama de flux pentru determinarea mediei și deviației ritmului de navigare a utilizatorului;
- fig. 7, diagrama de flux a determinării categoriilor probabile conform ritmului de navigare curent;
- fig. 8, diagrama de flux pentru determinarea probabilităților de a continua secvența începută cu diferite categorii.

În Fig. 1 este prezentată arhitectura metodei de predicție propuse, conform invenției. Astfel, modulul de procesare 4 va primi, pentru fiecare utilizator în parte, secvența curentă de navigare 1 ce conține paginile vizitate, momentele de timp ale vizitelor și categoriile acestor pagini. Momentul de timp al vizitelor ne va ajuta să evaluăm doi dintre factorii menționați anterior: momentul de timp la care se face navigarea și ritmul de navigare al utilizatorului, în timp ce paginile vizitate ne dau istoricul utilizatorului, cel de-al treilea factor amintit.

Datorită faptului că paginile web vizitate de utilizatori au o foarte mare diversitate, factor ce împiedică învățarea, vom reduce această diversitate prin înlocuirea paginilor web cu categoriile din care acestea fac parte, motiv pentru care această informație este de asemenea reținută în vederea procesării.

În plus, datorită faptului că o secvență tipică de navigare poate conține "cicluri" - un utilizator se poate întoarce la un anumit subiect (categorie de pagină) vizionat(ă) în trecut în cadrul aceleiași secvențe -, în timp ce pentru învățare avem nevoie de secvențe aciclice (încercăm să învățăm indicele unei anumite categorii într-o secvență și drept urmare acesta nu poate avea decât o singură valoare), acestea vor fi pre-procesate 3 pentru eliminarea ciclicității.

La modulul de procesare 4, pe lângă informațiile pre-procesate legate de secvența curentă de navigare, ajunge și istoricul utilizatorului 2, conținând secvențe anterioare de navigare.

Pe baza acestor informații, metoda identifică categoriile probabile spre a fi vizitate de către utilizator având în vedere momentul de timp actual **O1**, categoriile cel mai probabil să fie vizitate conform ritmului de navigare curent al acestuia **O2**, categoriile cel mai probabil să fie vizitate conform istoricului de navigare **O3** și predicția categoriei paginii următoare ce va fi vizitată de către client ținând cont de toți factorii considerați în analiză **O4**.

În continuare, se verifică dacă secvența de navigare este sau nu încheiată **5**. În cazul în care nu s-a detectat încheierea secvenței de navigare, întregul proces este reluat în vederea continuării analizei și predicției **10, 11, 12, 13**. În caz contrar, secvența trebuie salvată în istoricul utilizatorului. Pentru aceasta, mai întâi este determinat spațiul necesar stocării secvenței curente **6**. Datorită faptului că memoria asignată acestui procedeu poate fi limitată, înaintea salvării secvenței se face verificarea dacă există suficient spațiu disponibil sau nu **7**. Dacă există suficient spațiu disponibil, atunci secvența este adăugată la istoric **9** și metoda își încheie funcționarea. Dacă nu există suficient spațiu, atunci mai întâi sunt șterse o serie de intrări deja existente în istoric **8** pentru a se crea spațiul necesar, iar după aceea se stochează noua secvență în spațiul recent eliberat **9**. Ștergerea secvențelor are scopul pe de-o parte de a menține istoricul **2** în limite rezonabile din punctul de vedere al dimensiunii ocupate, iar pe de altă parte, de a furniza predicții cât mai apropiate de comportamentul actual al utilizatorului curent.

Având (pentru fiecare utilizator) secvențe de navigare curente **1**, în cadrul procesării **4** vom folosi câte un vector și o matrice independente pentru fiecare utilizator în parte (vectorul reprezentând secvența de navigare curentă **1**, iar matricea istoricul utilizatorului **2**). Matricea ce conține istoricul utilizatorului **2** va avea pe cele două dimensiuni următoarele informații:

- secvențele de pagini web vizitate de către utilizatorul respectiv în trecut și
- categoriile posibile pentru diferitele pagini web.

Informațiile salvate în matrice vor reprezenta indicii din secvență ai diferitelor categorii de pagini vizitate. Cu alte cuvinte, în matrice vom surprinde ordinea în care utilizatorul a vizitat diverse categorii de pagini. Datorită faptului că în astfel de matrici nu se poate salva decât o singură valoare pentru fiecare tuplu (în cazul nostru <secvență, categorie din secvență>) în timp ce o secvență tipică poate conține "cicluri" - un utilizator se poate întoarce la un anumit subiect (categorie de pagină) vizionat(ă) în trecut în cadrul aceleiași secvențe -, trebuie să avem grijă ca fiecare secvență din matrice să fie aciclică. Drept urmare, este nevoie de o pre-procesare a secvențelor de navigare inițiale **1** astfel încât să se poată salva toate informațiile asociate lor (chiar dacă există cicluri). Astfel, secvența de navigare curentă **1** este pre-procesată și transformată în unul sau mai mulți vectori de navigare curentă. Dacă se obțin mai mulți astfel de vectori, înseamnă că au fost identificate cicluri și în acest caz nu se menține în vederea procesării **4** decât ultimul vector. Ceilalți vectori vor fi salvați **9** în matricea reprezentând istoricul utilizatorului **2** dacă există suficient spațiu de stocare **7**, sau dacă nu, înainte de această operație, sunt eliminate **8** o serie de coloane din acea matrice de istoric **2** pentru a face loc pentru salvarea secvenței curente **9**. Odată finalizată procesarea **4** vectorului de navigare curentă, același lucru se întâmplă și cu acest vector: se adaugă **9** în matricea de istoric **2** dacă este posibil **7**, iar dacă nu, atunci mai întâi se elimină **8** niște coloane din istoric **2** și apoi este salvat vectorul respectiv **9**.

Algoritmul de pre-procesare 3, realizează, conform metodelor cunoscute, transformarea secvenței curente de navigare într-o serie de vectori aciclici, conținând categoriile paginilor vizitate (obținute de la un serviciu existent). De asemenea, se adaugă trei categorii fictive (start, continue, stop) în scopul delimitării sub-secvențelor.

În urma pre-procesării 3 se obține vectorul ce trebuie procesat. Detaliile procesării 4 sunt prezentate în Fig. 2. Procesarea 4 implică, pe de-o parte informațiile referitoare la vectorul curent de navigare, iar pe de alta, informațiile referitoare la istoricul de navigare al utilizatorului 2. Istoricul de navigare 2 este folosit pentru determinarea categoriilor probabile spre a fi vizitate de utilizator în funcție de diverse momente de timp 14, precum și pentru determinarea categoriilor probabile spre a fi vizitate de acesta în funcție de ritmul său de navigare 15. Odată obținute aceste informații, ele se combină cu vectorul de navigare curent pentru a determina care sunt categoriile probabile spre a fi vizitate la momentul curent de timp 10, respectiv categoriile probabile conform ritmului de navigare curent 11. De asemenea, vectorul de navigare curent și istoricul de navigare al utilizatorului 2 sunt folosite într-un modul de determinare a probabilității de continuarea a secvenței curente 12.

Odată obținute rezultatele de la modulele de determinare categorii probabile la momentul de timp curent 10 (O1), de determinare categorii probabile conform ritmului de navigare curent 11 (O2) și de determinare probabilitate de continuare a secvenței 12 (O3), acestea sunt combinate pentru a obține un singur set de valori pentru fiecare categorie în parte 16. În final, pe baza acestor valori se va face predicția categoriei următoarei pagini ce va fi vizitată de către utilizator O4 în cadrul modulului de predicție categorie 13.

În fig. 3 este detaliat modulul de determinare a categoriilor probabile în funcție de diferite momente de timp 14. În cadrul acestuia, mai întâi se construiește o matrice (M) având categoriile de pagini pe linie și sloturile de timp pe coloane 17 și se inițializează cu 0 fiecare intrare din matrice 18. În acest context, un slot de timp poate avea mai multe granularități, în funcție de cantitatea de date disponibilă și de necesitatea de predicție. În continuare este analizat istoricul utilizatorului 2 și pentru fiecare intrare în parte 19 se determină momentul de timp la care a fost accesată categoria respectivă 20, iar apoi se încadrează momentul respectiv de timp într-unul din slot-urile de timp definite 21. Pe baza acestei încadrări se va incrementa elementul din matrice (M_{ij}) corespunzător categoriei i și slotului de timp din care face parte aceasta (j) 22. Odată încheiată analiza întregului istoric de navigare al utilizatorului 2 se pot determina probabilitățile fiecărei categorii i pentru fiecare slot de timp j - $p(i|j)$ - (prin împărțirea lui M_{ij} la $\sum_j M_{ij}$) 23, precum și probabilitățile fiecărui slot de timp j - $p(j)$ - (prin împărțirea $\sum_i M_{ij}$ la $\sum_{ij} M_{ij}$) 24.

$$p(\text{categorie } i \mid \text{slot } j) = \frac{M_{ij}}{\sum_j M_{ij}}$$

$$p(\text{slot } j) = \frac{\sum_i M_{ij}}{\sum_{ij} M_{ij}}$$

Pe baza acestor probabilități se determină în continuare categoriile cele mai probabile să fie accesate la momentul actual de timp O1 în cadrul modulului de determinare categoriilor probabile la momentul curent de timp 10. Diagrama acestui proces este prezentată în fig. 4. Odată obținute probabilitățile fiecărei categorii i pentru fiecare slot de timp j 23, aceste valori

trebuie filtrate astfel încât să rămână numai acele valori specifice pentru slotul curent. Drept urmare, mai întâi se determină momentul de timp al ultimei categorii accesate **25** pe baza vectorului de navigare curent, după care se încadrează secvența curentă într-unul din sloturile de timp definite **26**. Având identificat slotul de timp curent, vom extrage din probabilitățile identificate pentru fiecare categorie i și pentru fiecare slot de timp j **23** doar acele probabilități care se întâlnesc în slotul curent de timp **10**. Aceste probabilități sunt furnizate drept ieșirea **01** a metodei.

În fig. 5 este prezentat în detaliu blocul de determinare a categoriilor probabile spre a fi vizitate în funcție de ritmul de navigare **15**. Pentru aceasta, mai întâi se construiește o matrice având pe linii categoriile de pagini disponibile, iar pe coloane ritmurile de navigare identificate. De exemplu, au fost considerate 4 ritmuri de navigare diferite, folosind media și abaterea standard (deviația) ale ritmului secvențelor anterioare ale utilizatorului calculate pe baza istoricului acestuia **2**. Totuși, în funcție de necesități, se poate defini un număr oricât de mare de astfel de ritmuri, folosind și alte metrici în afară de cele două sugerate mai sus. Cele 4 ritmuri de navigare considerate sunt:

1. Timp mediu mare între categorii, deviație mare → navigare cu pauze (Tip 1)
2. Timp mediu mare între categorii, deviație mică → navigare încetinită (Tip 2)
3. Timp mediu mic între categorii, deviație mare → navigare haotică (Tip 3)
4. Timp mediu mic între categorii, deviație mică → navigare accelerată (Tip 4)

Odată construită matricea cu categoriile de pagină pe linii și cele 4 ritmuri de navigare pe coloane, aceasta se inițializează cu valoarea 0 **28**. În continuare se determină media și deviația ritmului de navigare a utilizatorului pentru fiecare intrare din fiecare secvență din istoricul de navigare **29**. Acest modul va fi detaliat în fig. 6. Mai departe, pentru fiecare intrare din istoricul de navigare **30** trebuie determinat dacă intrarea respectivă este sau nu prima intrare din secvență **31**, deoarece pentru prima intrare nu se poate determina ritmul de navigare întrucât se consideră că nu există o intrare anterioară și atunci pentru aceasta nu avem cum să calculăm media sau deviația ritmului de navigare. Dacă este prima intrare **31 da**, atunci se trece la următoarea intrare (nu se calculează nimic) **32**. Altfel, **31 Nu**, se extrage media și deviația ritmului de navigare curente ale utilizatorului **33** calculate deja la pasul **29**. Odată obținute aceste valori, ele sunt comparate cu 2 praguri (unul pentru medie **34** și altul pentru deviație **35, 36**). Valorile celor 2 praguri pot fi determinate euristic, pot fi alese pseudo-aleator sau pot fi determinate pe baza istoricului utilizatorului **2**.

În cazul în care valorile sunt alese pseudo-aleator, se pot considera valori recunoscute ca fiind valide pentru diferite ritmuri de navigare (de exemplu, media timpului petrecut pe o pagină web - adică întârzierea dintre 2 pagini consecutive - este de 15 secunde iar deviația este de 3 secunde). Avantajul acestei metode este că nu necesită cunoștințe prea multe despre utilizator sau despre paginile vizitate și se poate aplica imediat ce un nou utilizator a fost înregistrat. Totuși, o astfel de metodă nu poate funcționa eficient pentru toată lumea și toate genurile de pagini web; ea trebuie adaptată diferitelor situații întâlnite în practică.

Dacă pragurile se determină pe baza istoricului utilizatorului **2**, atunci se pot calcula valorile mediei și deviației întârzierii de navigare și apoi aceste valori să fie considerate drept praguri. Avantajul acestei metode este că valorile sunt adaptate foarte bine utilizatorului.

În funcție de valorile celor 2 praguri și de comparațiile mediei **34** și deviației **35, 36** cu aceste praguri, se stabilește ritmul curent de navigare ca făcând parte din unul din cele 4 tipuri și apoi se incrementează valoarea corespunzătoare categoriei curente și ritmului curent de navigare **37, 38, 39, 40**. Mai departe, se verifică dacă s-au terminat de prelucrat toate intrările **41**. Dacă mai sunt intrări de prelucrat **41 Nu**, atunci se trece la următoarea intrare **32** și procesul se reia. În caz contrar **41 Da** se pot calcula probabilitățile fiecărei categorii i pentru fiecare ritm de navigare j ($p(i|j)$) prin împărțirea valorii lui M_{ij} la $\sum_j M_{ij}$ **42**. De asemenea, tot aici se pot determina probabilitățile fiecărui ritm de navigare j ($p(j)$) prin împărțirea $\sum_i M_{ij}$ la $\sum_{i,j} M_{ij}$.

$$p(\text{categorie } i \mid \text{ritm de browsing } j) = \frac{M_{ij}}{\sum_j M_{ij}}$$

$$p(\text{ritm de browsing } j) = \frac{\sum_i M_{ij}}{\sum_{i,j} M_{ij}}$$

În fig. 6, așa cum s-a menționat deja, este descrisă modalitatea de calcul a mediei și deviației ritmului de navigare curente ale utilizatorului **29**. Aceste calcule se pot realiza considerând întreaga secvență de navigare a utilizatorului sau numai ultima parte a acestei secvențe, având în vedere faptul că ritmul de navigare al utilizatorului se poate modifica pe parcursul unei secvențe de navigare în funcție de diverși factori, cum ar fi tipul de informație căutată sau posibila plictiseală a utilizatorului ce poate interveni după un anumit timp. Drept urmare, înainte de a începe calculul, se stabilește cantitatea de informații referitoare la navigarea anterioară folosite pentru aceste calcule, adică fereastra din care extragem informațiile legate de vizitele anterioare ale utilizatorului din secvența curentă. Această fereastră poate fi definită fie ca întreaga secvență curentă a utilizatorului (nemaiținând cont de modificările de ritm ce pot apare în interiorul secvenței), fie ca cea mai recentă parte a acestei secvențe, ce poate fi definită la rândul ei fie ca interval de timp (de exemplu, vizitele din ultimele 30 de secunde, ultimele 5 minute sau ultima oră față de momentul curent), fie ca număr de categorii de pagini vizitate (de exemplu, ultimele 3, 5, 10 categorii de pagini vizitate care fac parte din secvența curentă), fie ca o combinație a acestor doi factori (număr de categorii de pagini vizitate dintr-un anumit interval de timp: de exemplu, ultimele 10 categorii vizitate în ultimele 5 minute). În fig. 6 este detaliată varianta ce ține cont de un număr fix de categorii de pagini vizitate (dimensiunea ferestrei, `window_size`) pentru a evidenția ce probleme ridică o astfel de abordare, celelalte variante reprezentând simple adaptări ale acestei abordări. Astfel, pentru prima secvență j ($= 0$) din istoricul de navigare **43**, odată stabilită dimensiunea ferestrei (`window_size`), se vor crea doi vectori având un număr de componente egal cu dimensiunea ferestrei (`window_size`): t și δ **44**. Vectorul t reprezintă vectorul timpilor de accesare a diferitelor categorii de pagini, în timp ce vectorul δ reprezintă diferența de timp dintre accesarea curentă și cea anterioară. De asemenea, tot în acest pas se vor construi alți doi vectori m și dev având un număr de componente egal cu dimensiunea secvenței. M reprezintă vectorul mediilor diferențelor de timp dinaintea fiecărei intrări din secvența j (ținând cont de dimensiunea ferestrei și de intrările care se încadrează în această fereastră), iar dev reprezintă deviațiile standard ale diferențelor de timp calculate pentru aceleași intrări. Dev se va calcula pe baza vectorilor m și δ , conform formulei:

$$dev_i = \sqrt{\frac{\sum_k (\delta_k - m_i)^2}{window_size - 1}}$$

unde k reprezintă indicii intrărilor ce fac parte din fereastra aleasă și variază de la 0 la $window_size - 1$.

În afară de acești vectori, algoritmul de determinare a mediei și deviației de navigare curente ale utilizatorului **29** mai folosește trei variabile: s , reprezintă suma diferențelor de timp din fereastră și va fi folosită pentru a calcula media diferențelor de timp din fereastră: $m_i = s/window_size$; i , reprezintă indicele elementului din secvență și este folosit pentru trecerea de la o intrare la alta și j care reprezintă indicele secvenței curente din istoricul utilizatorului **2**.

După construcția celor patru vectori (m , dev , t și δ) **44**, aceștia sunt inițializați cu valoarea 0 și la fel se întâmplă și cu valorile lui s și i **45** (j a fost deja inițializat la pasul **43**). În continuare, pentru fiecare intrare i **46**, se determină momentul de timp la care a fost accesată categoria respectivă și se salvează această valoare în $t[i\%window_size]$ **47** unde operatorul $\%$ reprezintă operația de obținere a restului împărțirii. Cu alte cuvinte, cât timp valoarea lui i este mai mică decât dimensiunea ferestrei, momentele de timp se salvează chiar în ordinea în care au fost citite. Odată ajunsă valoarea lui i la valoarea dimensiunii ferestrei ($i = window_size$), înseamnă că cea mai veche valoare din fereastră trebuie ștearsă pentru că intrarea respectivă nu mai face parte din fereastră (indicii din fereastră sunt cuprinși între 0 și $window_size - 1$ și atunci când se găsește intrarea cu indicele $window_size$, înseamnă că cele mai recente $window_size$ intrări sunt intrările din secvență cu valorile cuprinse între 1 și $window_size$). Drept urmare, informațiile aferente intrării curente trebuie să suprascrie datele referitoare la cea mai veche intrare din fereastră (intrarea cu indicele 0), iar acest lucru se realizează folosind operatorul modulo ($\%$), întrucât restul împărțirii lui $window_size$ (indicele curent din secvență) la $window_size$ (dimensiunea ferestrei), este 0 și astfel informațiile referitoare la intrarea curentă vor suprascrie informațiile existente deja pe poziția 0 din vectorul t .

Odată salvat momentul de timp al intrării curente, se verifică dacă această intrarea este prima din secvență **48**, întrucât dacă este prima intrare nu se pot determina metricile necesare pentru ea. Dacă este prima intrare **48 Da**, atunci se trece la următoarea intrare prin incrementarea indicelui i **53**. Altfel **48 Nu**, se verifică dacă valoarea lui i este sau nu mai mare decât $window_size$ **49**. Această verificare se face cu scopul optimizării calculului valorii s (suma diferențelor de timp din fereastră). Astfel, dacă valoarea lui i este mai mare decât valoarea $window_size$ **49 Da** vor trebui actualizate niște valori deja salvate cu noile valori calculate ca urmare a intrării curente. Una din aceste valori este s (suma diferențelor de timp din fereastră). Din valoarea lui s va trebui să se scadă prima valoare a întârzierii și în schimb să se adauge noua valoare calculată pentru intrarea curentă. Valoarea veche a întârzierii este eliminată folosind formula $s = s - \delta[i\%window_size]$ **50**. Mai departe, trebuie adăugată noua întârziere, dar pentru a putea face acest lucru, mai întâi trebuie calculată și salvată această nouă valoare, folosind formula $\delta[i\%window_size] = t[i\%window_size] - t[(i-1)\%window_size]$ **51**. Odată avută valoarea noii întârzieri, ea se adaugă la suma diferențelor de timp din fereastră (s): $s = s + \delta[i\%window_size]$ **51**. În acest moment se pot calcula

media și deviația ritmului de navigare curente folosind formulele de mai sus. Aceste valori se vor salva pe poziția i în vectorii m și dev .

Dacă în schimb valoarea lui i nu este mai mare decât valoarea `window_size` **49 Nu**, înseamnă că nu va trebui să actualizăm valori, ci doar să adăugăm noile informații care fac parte din fereastra. Drept urmare se va sări peste pasul **50** și se va trece direct la pasul **51**, în care se calculează și se salvează noua întârziere ca diferență de timp între intrarea curentă și cea anterioară, iar după aceea se calculează valoarea lui s și pe baza acesteia, valorile mediei și deviației de navigare de la momentul curent de timp. Având în vedere că valorile lui i sunt mai mici decât `window_size`, modulul (%) nu influențează operațiile realizate în pasul **51**, astfel încât se pot trata unitar și eficient atât cazurile când i este mai mare decât `window_size` cât și atunci când i este mai mic.

Odată calculate valorile mediei și deviației de navigare curente se poate trece la următoarea intrare **52** pentru a repeta procesul. Pentru aceasta se verifică dacă intrarea curentă a fost sau nu ultima din secvența curentă **53**. Dacă nu **53 Nu**, atunci procesul se reia de la pasul **46**. Altfel **53 Da**, procesarea secvenței curente s-a încheiat și se încearcă trecerea la următoarea secvență de prelucrat. În următorul pas **54** se verifică dacă secvența a cărei procesare tocmai s-a încheiat a fost ultima din istoricul utilizatorului **2**. Dacă da **54 Da**, atunci procesarea se încheie întrucât întreaga istoric al utilizatorului **2** a fost procesat. În caz contrar se trece la următoarea secvență **55** și apoi procesul se reia de la pasul **44**.

În fig. 7 este prezentat modul în care se decid probabilitățile fiecărei categorii de pagini în parte, astfel încât să se țină cont de ritmul curent de navigare **O2**. Astfel, odată determinate probabilitățile fiecărei categorii i pentru fiecare ritm de navigare **42**, acestea sunt filtrate în funcție de ritmul de navigare curent **56**, pentru a returna numai acele probabilități care sunt adaptate acestui ritm **11 (O2)**. Determinarea ritmului de navigare curent **56** se face cu ajutorul modulului de determinare a mediei și deviației de navigare a utilizatorului **29**, cu observația că în acest caz nu avem decât o singură secvență (deci $j = 0$ mereu) și în plus nu ne interesează decât valorile ultimelor valori ale mediei și deviației de navigare a utilizatorului.

În fig. 8 este descrisă modalitatea de calcul a categoriilor cel mai probabil să fie vizitate conform istoricului de navigare **O3**, din cadrul modulului de determinare a probabilității de continuare a secvenței **12**, folosind istoricul de navigare al utilizatorului **2**. Așa cum am menționat deja, informațiile legate de istoricul de navigare al utilizatorului sunt salvate într-o matrice, astfel încât să se surprindă ordinea în care utilizatorul a vizitat diverse categorii de pagini. Drept urmare, informațiile salvate în această matrice vor reprezenta indicii din secvență ai diferitelor categorii de pagini vizitate. Folosind această matrice **2** și secvența parțială de navigare **1**, se va determina probabilitatea fiecărei categorii de pagini de a fi vizitată în continuare cu ajutorul unui procedeu de recomandare. Cu alte cuvinte, se folosește un procedeu de recomandare care pe baza unei secvențe parțiale existente la un moment dat și a istoriei cunoscute despre un utilizator (secvențe anterioare de navigare ale acestuia), să învețe ordinea în care acesta vizitează diferite categorii de pagini și apoi să sugereze posibilități de continuare a unor noi secvențe (parțiale) de navigare. Procedeele de recomandare se bazează pe realizarea unei matrici utilizator - obiect, unde sunt salvate, de obicei, ratingurile date de către fiecare utilizator obiectelor folosite. Această matrice este de obicei rară, întrucât numărul mare de obiecte și utilizatori face practic imposibilă existența

unui rating pentru fiecare tuplu <utilizator,obiect> (care presupune ca fiecare utilizator să fi folosit și evaluat fiecare dintre obiectele din matrice). Varianta clasică a unui astfel de procedeu de recomandare a fost adaptată astfel încât să țină cont de nevoile / informațiile disponibile în invenția noastră. Astfel, în loc să avem mai mulți utilizatori, noi vom avea mai multe secvențe de navigare, iar pe post de obiecte vom avea categoriile de pagină identificate; în cazul nostru, în locul tuplurilor clasice <utilizator, obiect> vom avea tupluri <secvență, categorie din secvență>. Cu alte cuvinte, vom folosi matricea istoricului utilizatorului **2** având pe cele două dimensiuni secvențele de pagini web vizitate de către utilizatorul respectiv în trecut și categoriile posibile pentru diferitele pagini web. Folosind această matrice, ceea ce învață procedeul de recomandare este diferit față de versiunea clasică (unde detecta un rating pentru un anumit obiect). Astfel, procedeul de recomandare învață de fapt ordinea în care sunt parcurse diferite categorii de pagini, putând fi folosit ulterior pentru sugerarea ordinii în care vor fi vizitate diverse categorii pentru secvențe noi, nefinalizate.

Acest modul preia vectorul de navigare curent (care este incomplet) și istoricul de navigare al utilizatorului **2** și îl trimite unui procedeu de recomandare **57**. Procedeul de recomandare poate fi orice procedeu de recomandare clasic sau îmbunătățit pe baza noilor descoperiri din domeniu, adaptările făcute anterior neimpunând nicio restricție asupra funcționalității acestuia, făcând posibilă folosirea oricărui astfel de procedeu. În continuare, procedeul de recomandare va furniza câte o valoare pentru fiecare categorie nevizitată deja (care nu face parte din secvența parțială de navigare) **58**, inclusiv pentru categoriile continue și stop introduse suplimentar în cadrul algoritmului de pre-procesare. Această valoare reprezintă ordinea probabilă a fiecărei categorii analizate în secvența incompletă dată spre evaluare. Știind câte categorii au fost deja vizitate în cadrul secvenței curente incomplete, se poate determina, pentru fiecare categorie rămasă nevizitată, diferența dintre următorul indice al secvenței (lungimea secvenței + 1) - indicele pe care ar urma să fie vizitată o nouă categorie - și valoarea prezisă de procedeul de recomandare. Această diferență este transformată într-o probabilitate de vizitare **59** prin împărțirea ei la numărul total de categorii + 1 și scăderea valorii obținute din valoarea 1. Acest lucru se întâmplă datorită faptului că cea mai lungă secvență ce se poate construi poate avea maxim numărul total de categorii +2 elemente (start și continue sau stop): start (începutul secvenței), toate categoriile (fiecare o singură dată pentru că altfel am avea ciclu și secvența s-ar sparge în două sub-secvențe) și apoi ori continue (s-a găsit ciclu) ori stop (s-a terminat secvența de navigare), iar diferența maximă de indici din această secvență este numărul total de categorii + 1. Astfel, folosind această metodologie se poate determina probabilitatea ca o categorie de pagini nevizitată încă să fie următoarea pagină vizitată. Totuși problema estimării probabilităților pentru toate categoriile de pagini nu este încă rezolvată, deoarece mai sunt și categoriile de pagini deja existente în secvență (vizitate), a căror probabilitate de re-vizitare trebuie estimată **60**. Pentru a evalua aceste probabilități se folosesc două informații: probabilitatea categoriei continue (probabilitatea de a se încheia un ciclu la următorul pas) și probabilitățile categoriilor deja vizitate (cu excepția categoriei Start) de a se afla pe prima poziție în istoricul utilizatorului **2**. Cu alte cuvinte, mai întâi estimăm care este probabilitatea de a încheia ciclul în pasul următor, după care trebuie să evaluăm care este categoria cea mai probabilă să încheie ciclul din cele vizitate deja. Cum prima informație este furnizată de către procedeul de recomandare, ne mai rămâne să evaluăm cea de-a doua informație. Datorită algoritmului de

pre-procesare, de fiecare dată când se identifică un nou ciclu în secvența de navigare, aceasta este spartă în două subsecvențe, cea de-a doua începând cu categoria care a dus la crearea ciclului. Drept urmare, se poate evalua probabilitatea fiecărei categorii de a genera cicluri prin determinarea numărului de situații în care acest lucru s-a întâmplat și împărțirea acestui număr la numărul total de cicluri identificate în toată istoria de navigare a utilizatorului **2**. Având aceste probabilități, ele se filtrează astfel încât să rămână numai probabilitățile acelor categorii care au fost deja vizitate în cadrul secvenței, după care valorile obținute se înmulțesc cu probabilitatea de a încheia ciclul (determinată pe baza rezultatelor procedurii de recomandare). Înmulțirea celor două probabilități are loc datorită faptului că ambele evenimente (aparitia ciclului și categoria care să creeze ciclul respectiv) trebuie să apară concomitent și astfel probabilitățile lor de apariție trebuie combinate (prin înmulțire). După acest ultim pas avem disponibile toate probabilitățile de continuare a secvenței curente **O3** și acestea se pot furniza la ieșirea modulului **61**.

Odată obținute valorile probabilităților fiecărei categorii în parte în funcție de factorii considerați (momentul actual de timp, modulul **10**, ieșirea **O1**, ritmul curent de navigare, modulul **11**, ieșirea **O2** și istoricul utilizatorului, modulul **12**, ieșirea **O3**), aceste valori trebuie combinate astfel încât procedeul să furnizeze o singură probabilitate finală pentru fiecare categorie în parte **16**. Practic, procedeul trebuie să furnizeze valoarea probabilității fiecărei categorii i dându-se momentul actual de timp, ritmul curent de navigare și istoricul utilizatorului: $p(\text{categorie}_i \mid \text{momentul actual de timp, ritmul curent de navigare, istoricul utilizatorului})$. Combinarea probabilităților se poate face fie prin înmulțire directă a probabilităților corespunzătoare aceleiași categorii, fie prin adunarea acestora, fie prin adunarea logaritmului acestora, fie prin utilizarea unei sume ponderate astfel încât să se poată da o importanță diferită fiecărei din cele trei categorii de probabilități considerate. Dacă cele trei probabilități se înmulțesc direct, se consideră că cei trei factori (momentul actual de timp, ritmul curent de navigare și istoricul utilizatorului) sunt independenți și atunci probabilitățile implicate de fiecare dintre aceștia factorizează, rezultând:

$$p(\text{categorie}_i \mid \text{momentul actual de timp, ritmul curent de navigare, istoricul utilizatorului}) = p(\text{categorie}_i \mid \text{momentul actual de timp}) * p(\text{categorie}_i \mid \text{ritmul curent de navigare}) * p(\text{categorie}_i \mid \text{istoricul utilizatorului}).$$

Datorită faptului că se lucrează cu probabilități având valori cuprinse între 0 și 1, unele dintre ele putând fi extrem de mici, este posibil ca prin înmulțirea lor să se ajungă foarte repede la valoarea 0 (din cauza depășirii preciziei calculatoarelor). De aceea, în unele cazuri se poate apela la un artificiu: logaritmare în prealabil a formulei de mai sus. Astfel, produsul se transformă în sumă și drept urmare se poate face o însumare a logaritmului probabilităților (în loc de o înmulțire directă a acestora) pentru a evita aceste probleme. Uneori, pentru a mai simplifica calculele, în loc să se însumeze logaritmul probabilităților, se însumează chiar valorile probabilităților, chiar dacă rezultatele nu sunt la fel de corecte. De asemenea, folosind suma (logaritmului) probabilităților în locul produsului acestora, se poate schimba importanța dată fiecărui factor în parte prin adăugarea câte unui coeficient de importanță fiecărui factor în parte. Acești coeficienți nu apar întâmplător, prezența lor fiind explicată de posibilele interacțiuni existente între diverși factori, interacțiuni ce nu pot fi surprinse de modelul inițial în care s-a presupus independența factorilor prezentați. Drept urmare, dacă

există influențe între factori, acei factori cu influențe mai mari vor avea coeficienți mai mari în cadrul sumei. De asemenea, dacă nu se dorește să se țină cont de vreun factor, acesta poate primi coeficientul 0 și astfel predicția lui este ignorată.

Ultimul pas al aplicației, predicție categorie **13**, are rolul de a determina și furniza la ieșire cea mai probabilă categorie dându-se istoricul utilizatorului, factorii de care se ține cont (momentul actual de timp, ritmul curent de navigare și istoricul utilizatorului), precum și eventualii coeficienți de importanță acordați fiecăruia dintre acești factori. În acest pas sunt comparate valorile probabilităților tuturor categoriilor (obținute de la modulul de combinare a probabilităților **16**) și se va furniza la ieșire acea categorie care are probabilitatea maximă **O4**.

Exemplul 1 de aplicare a metodei propuse.

Într-un exemplu nelimitativ de aplicare a invenției, metoda propusă poate fi folosită pentru cazul în care navigarea se face de pe un sistem desktop (calculator personal). În acest caz, metoda va fi implementată chiar pe calculatorul în cauză și va analiza istoria de navigare a utilizatorului, făcând predicții legate de următoarea/următoarele categorii de pagini ce vor fi vizitate de acesta. Pentru aceasta, metoda va fi implementată în browser, de exemplu într-o extensie de browser (scrisă în JavaScript).

Datele de intrare sunt reprezentate de o secvență de navigare de forma:

```
2015-03-02 11:29:11 impbeacon {"location":"www.google.es", "impId":"07be8c3b-302f-457e-864f-ab4953100636", "vertical":"Internet_and_Telecom/Search_Engine", "pk":0, "appInstanceId": "3AA812EC-A599-438B-9664-B747296AB527"} ES
```

.....

```
2015-03-02 13:49:33 impbeacon {"location":"www.angrybirdsgames.com", "impId":"8ebad827-881a-4f8d-8104-38164fdd1689", "vertical":"Games", "pk":0, "appInstanceId": "3AA812EC-A599-438B-9664-B747296AB527"} ES
```

Procedeul va determina probabilitățile fiecărei categorii în parte pentru fiecare slot de timp **14**. De exemplu, dacă se consideră 3 sloturi de timp (dimineața, prânz, seara), atunci se vor obține probabilitățile din Tabelul 1 pentru 5 categorii de pagini considerate (Internet_and_Telecom, Games, Adult, Finance și Other) + încă două suplimentare (Continue și Stop).

Tabel 1. Probabilitățile fiecărei categorii în parte pentru fiecare slot de timp

Categorie \ Slot de timp	Internet_and_Telecom	Games	Adult	Finance	Other	Continue	Stop
Dimineața	0.46	0.23	0	0.3	0	0.31	0.06
Prânz	0.38	0.08	0	0.15	0.38	0.44	0.23
Seara	0.23	0.3	0.15	0	0.3	0.25	0.69

De asemenea, se vor determina probabilitățile fiecărei categorii în parte pentru fiecare ritm de browsing al utilizatorului **15**, obținându-se valorile din Tabelul 2.

În fine, dându-se o secvență parțială conținând categoriile Start, Internet_and_Telecom și Games, folosind procedeul de recomandare **12**, se vor determina indicii următori pentru celelalte categorii rămase: Adult - 5, Finance - 4, Other - 3, Continue - 3, Stop - 5.

Probabilitățile obținute pentru fiecare din aceste categorii sunt Adult - 0.66, Finance -0.83 , Other - 1, Continue - 1, Stop - 0.66. De asemenea, pentru cele 2 categorii care pot să determine ciclul se obțin probabilitățile Internet_and_Telecom - 0.8 și Games - 0.2. Atunci ieșirea **O3** va fi: Internet_and_Telecom - 0.8, Games - 0.2, Adult - 0.66, Finance -0.83 , Other - 1, Continue - 1, Stop - 0.66.

Tabel 2. Probabilitățile fiecărei categorii în parte pentru fiecare slot de timp

Categorie Ritm navigare	Internet_and_Telecom	Games	Adult	Finance	Other	Continue	Stop
Tip 1	0.17	0.67	0	0	0.17	0.13	0.38
Tip 2	0.43	0.14	0	0.14	0.29	0.31	0.31
Tip 3	0.31	0.15	0	0.23	0.31	0.38	0.15
Tip 4	0.46	0.08	0.15	0.15	0.15	0.19	0.15

Știind că navigarea curentă se face în slotul de la prânz, ieșirea **O1** a procedurii va fi reprezentată de valorile Internet_and_Telecom - 0.38, Games - 0.08, Adult - 0, Finance -0.15, Other - 0.38, Continue - 0.44, Stop - 0.23. Dacă ritmul de navigare al utilizatorului este de tipul T2, ieșirea **O2** a procedurii va consta în valorile: Internet_and_Telecom - 0.43, Games - 0.14, Adult - 0, Finance -0.14, Other - 0.29, Continue - 0.31, Stop - 0.31.

Pentru a obține probabilitățile finale, trebuie înmulțite valorile reprezentând ieșirile **O1**, **O2** și **O3** și apoi normalizate valorile obținute. Drept urmare, probabilitățile finale vor fi: Internet_and_Telecom - 0.294, Games - 0.005, Adult - 0, Finance - 0.039, Other - 0.248, Continue - 0.307, Stop - 0.106. Pe baza acestor valori, se observă că vom obține valoarea maximă pentru categoria Continue (30.7%), ceea ce înseamnă că cel mai probabil utilizatorul se va întoarce la una din cele două categorii vizitate deja. Dintre cele două categorii, cea care are probabilitatea mai mare (apropiată de cea a categoriei Continue) este categoria Internet_and_Telecom (29.4%) și drept urmare se va alege drept categoria următoarei pagini ce va fi vizitată de către utilizator, aceasta reprezentând ieșirea **O4**.

Exemplul 2 de aplicare a metodei propuse. Într-un alt exemplu nelimitativ de realizare a invenției, aplicația poate fi folosită pe un dispozitiv mobil (telefon sau tabletă). Față de varianta de pe calculatorul personal, acum datele nu mai sunt procesate local, datorită resurselor limitate de care dispune dispozitivul mobil, ci sunt trimise la un server unde se fac prelucrările prezentate în exemplul anterior. Browserul dispozitivului mobil va colecta informațiile de navigare (aceleași ca în exemplul anterior) și apoi le va trimite către un server, de exemplu folosind formatul Java Script Object Notation, JSON. Serverul poate fi implementat pe o singură mașină, caz în care mecanismul de transmisie de pe server folosește, de exemplu, limbajul PHP și un server web (de exemplu, Tomcat), sau poate fi implementat într-un cloud, în cazul în care se decide să se rețină mai multe date pentru predicție (și unde implicit este nevoie de o mai mare putere de calcul pentru procesarea acestora). În acest caz modulul de transmisie de pe server este implementat, de exemplu, în Scala. La sfârșitul procesărilor, categoria cea mai probabilă (de exemplu, ca la Exemplul 1 Internet_and_Telecom) e returnată la dispozitivul mobil, unde e folosită în conjuncție cu un alt modul de pre-încărcare a paginilor din categoria prezisă.

Revendicări

1. Metodă de predicție individualizată a categoriilor paginilor de internet ce urmează să fie vizitate de un utilizator, destinată utilizării pe calculatoare personale și dispozitive mobile (telefoane, tablete), în scopul accelerării și simplificării navigației, prin oferirea, la momentul potrivit de timp, de sugestii pertinente de pagini web, ușor de accesat de către utilizator, **caracterizată prin aceea că** are următoarele faze:
 - a. Pre-procesare **3**, în care, pornind de la secvența curentă de navigare a utilizatorului și de la istoricul său de navigare se crează unul sau mai mulți vectori aciclici cu categoriile paginilor vizitate, în ordinea vizităților;
 - b. Procesare **4**, în care se prezice categoria următoarei pagini de internet ce va fi accesată de utilizator. Această predicție se face combinând două tehnici diferite de predicție: una bazată pe folosirea unui procedeu de recomandare **12**, iar cealaltă, o metodă statistică bazată pe compararea probabilităților diferitelor categorii de pagini de a fi accesate **10, 11**; și folosind trei factori diferiți pentru realizarea predicției: istoria navigării utilizatorului, momentul actual de timp și ritmul curent de navigare al utilizatorului;
 - c. Adăugarea informației pre-procesate în istoricul de navigare al utilizatorului.
2. Metoda de predicție, conform revendicării 1, **caracterizată prin aceea că** ritmul curent de navigare al utilizatorului este determinat, în cadrul modulului de procesare **4**, de către metoda de determinare a categoriei probabile conform ritmului de navigare **11**, pe baza mediei și deviației vitezei de navigare, ca aparținând uneia din următoarele 4 categorii: navigare cu pauze, navigare încetinită, navigare haotică și navigare accelerată.
3. Metoda de predicție bazată pe un procedeu de recomandare, conform revendicării 1, **caracterizată prin aceea că**, în cadrul modulului de procesare **4**, metoda de determinare al probabilității de continuare a secvenței **12** ține cont de întreaga istorie de navigare a utilizatorului și poate genera predicții pentru o întreagă secvență de navigare (nu doar pentru următoarea categorie ce va fi accesată) împreună cu metoda de pre-procesare a datelor **3** astfel încât să poată fi folosite într-un astfel de procedeu de recomandare.
4. Metoda de predicție, conform revendicării 3, **caracterizată prin aceea că** metoda de determinare al probabilității de continuare a secvenței **12** folosește transformarea rezultatelor furnizate de procedeu de recomandare în probabilități:
 - o pentru categoriile nevizitate, prin estimarea diferenței dintre ordinea furnizată de procedeu de recomandare și numărul de ordine al categoriei următoare ce ar trebui vizitate în cadrul secvenței parțiale, și normalizarea acestei diferențe prin împărțirea la diferența maximă ce poate apărea (numărul total de categorii + 1);
 - o pentru categoriile re-vizitate, prin înmulțirea probabilității categoriei Continue (determinată folosind metodologia de mai sus) cu probabilitatea fiecărei categorii vizitate deja de a genera cicluri.

21-08-2015

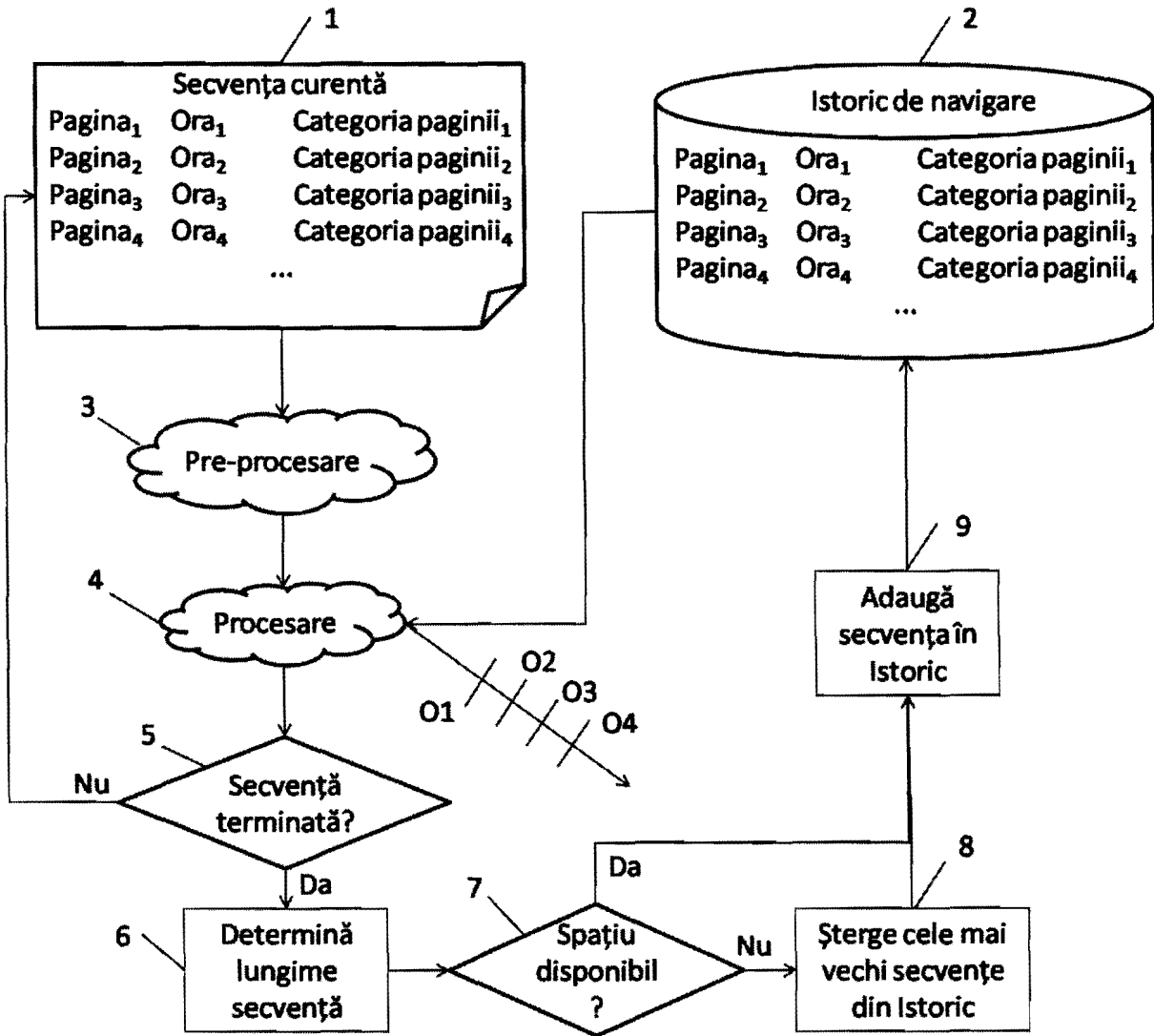


Fig.1

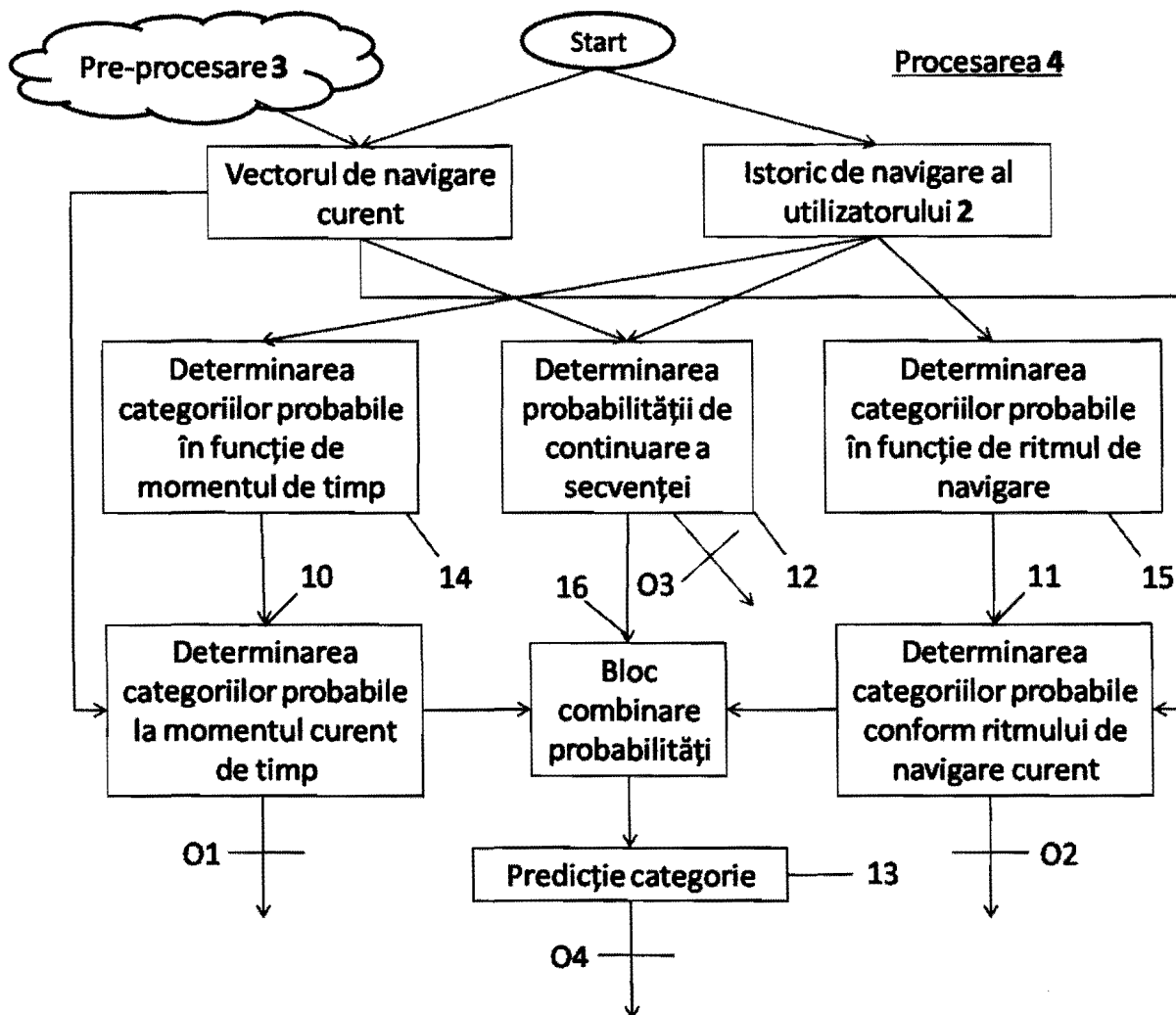


Fig. 2

21-08-2015

**Determinarea
categoriilor
probabile în funcție
de momentul de
timp 14**

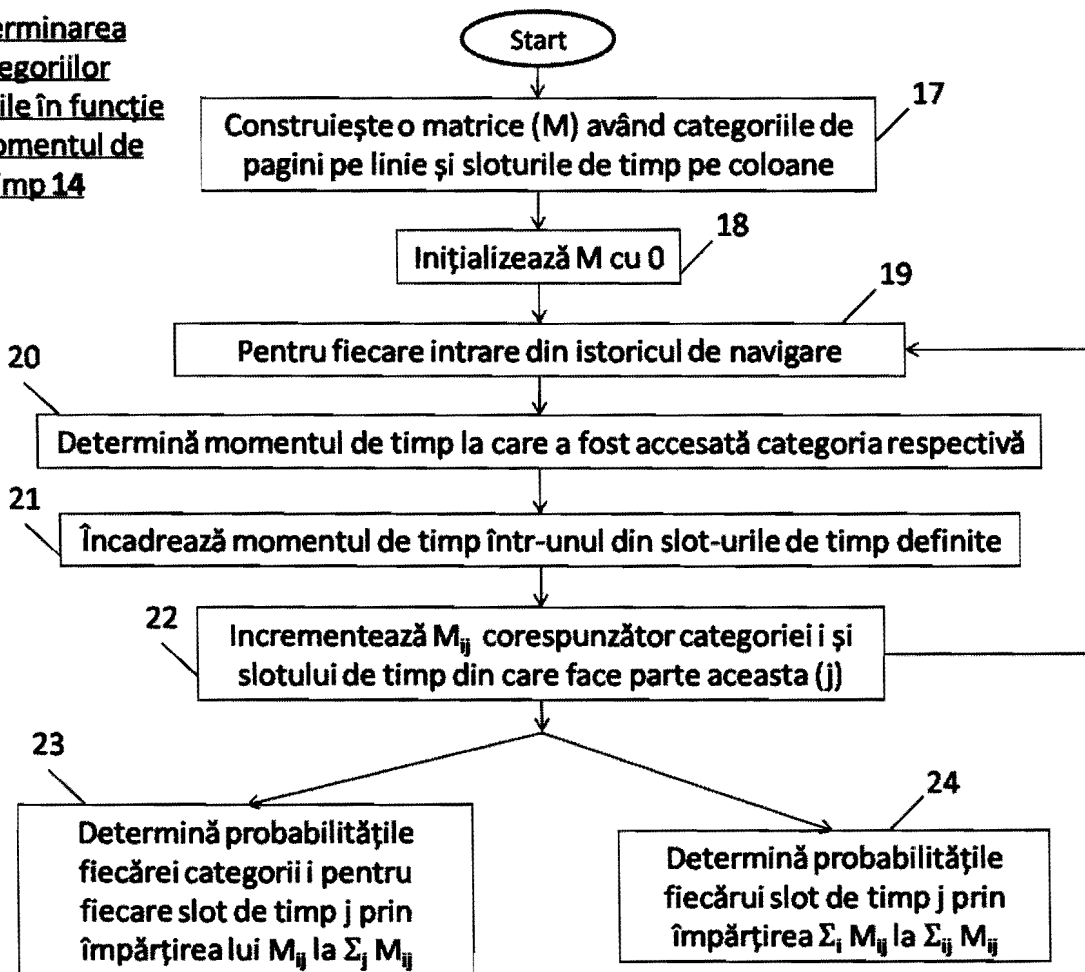


Fig. 3

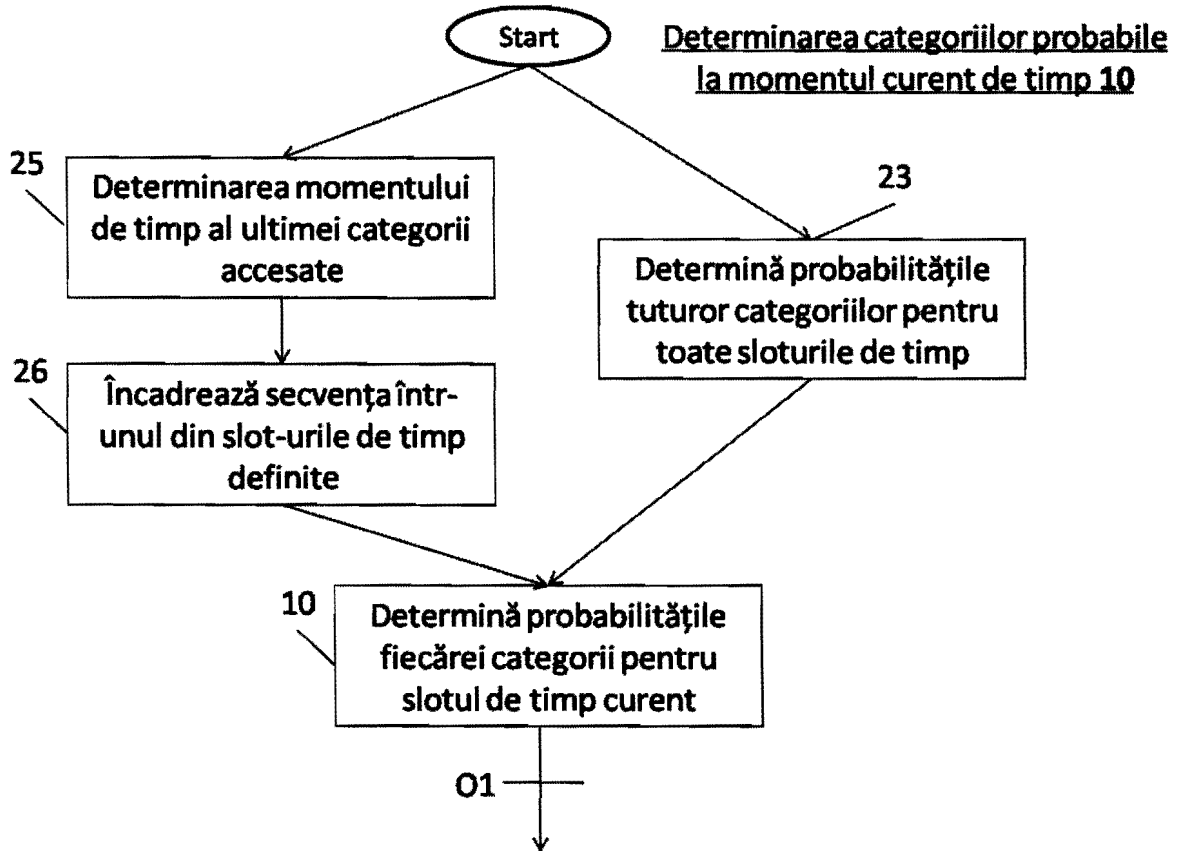


Fig. 4

Determinarea categoriilor probabile în funcție de ritmul de navigare 15

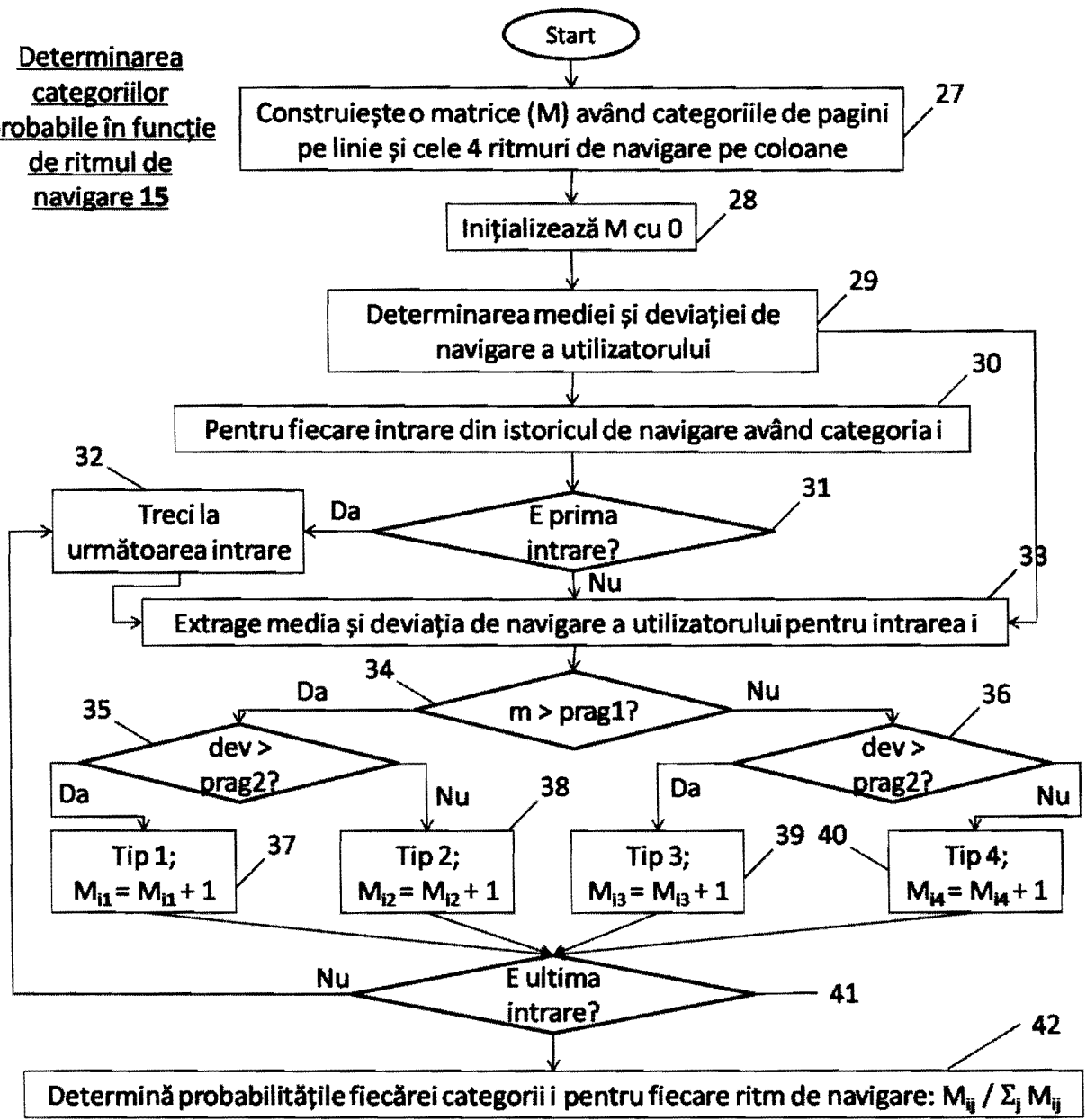


Fig. 5

**Determinarea mediei și deviației
ritmului de navigare a utilizatorului 29**

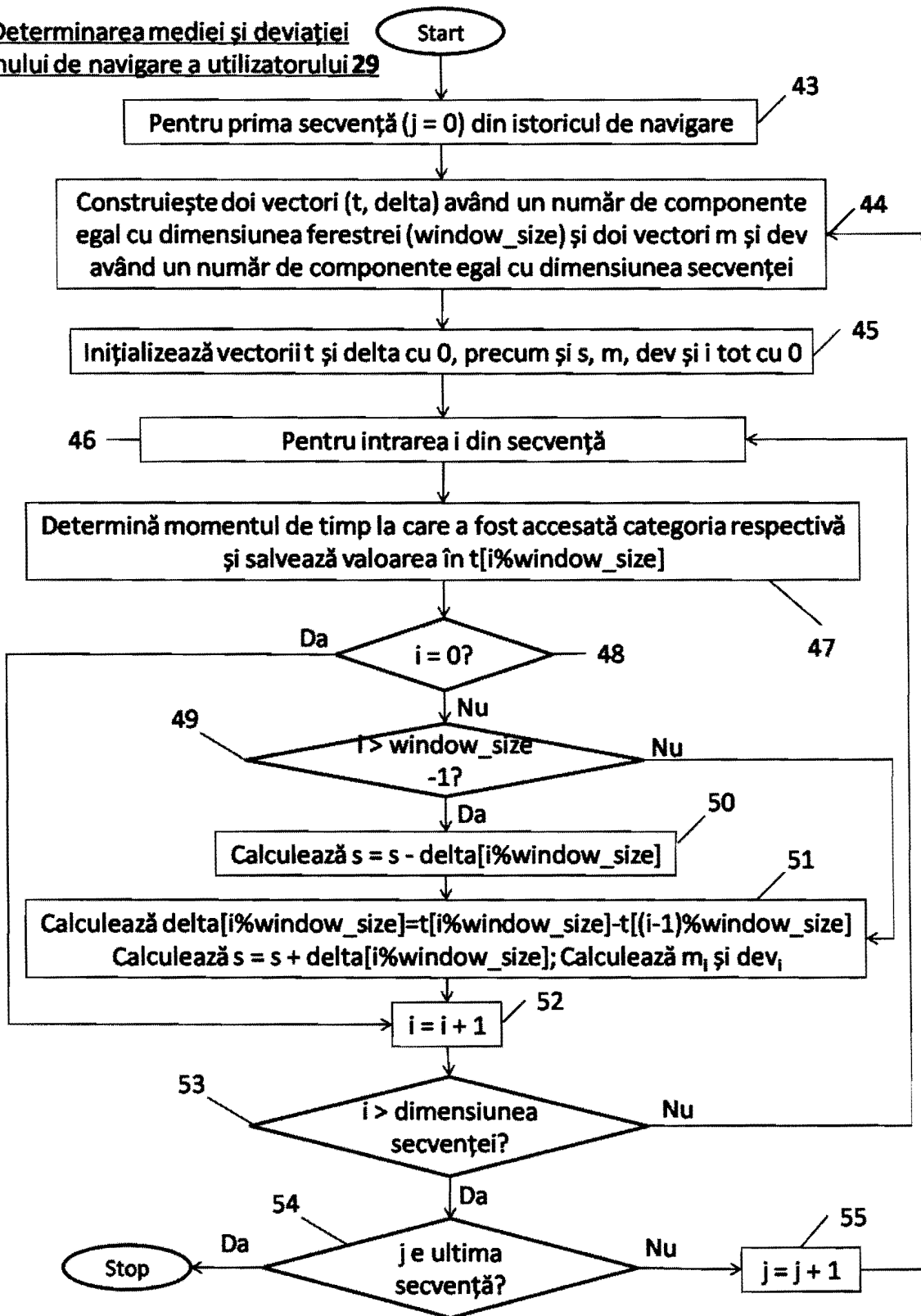
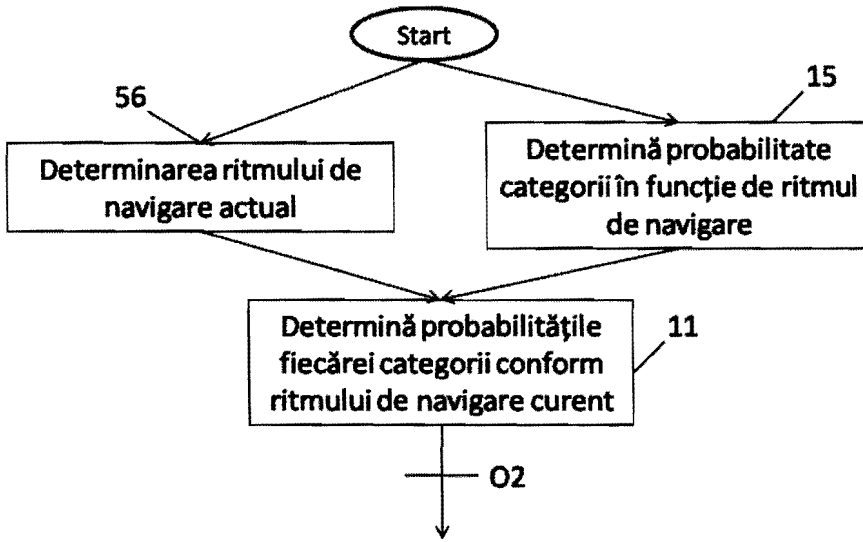


Fig. 6

21 -08- 2015



Determinarea categoriilor probabile conform ritmului de navigare curent 11

Fig. 7

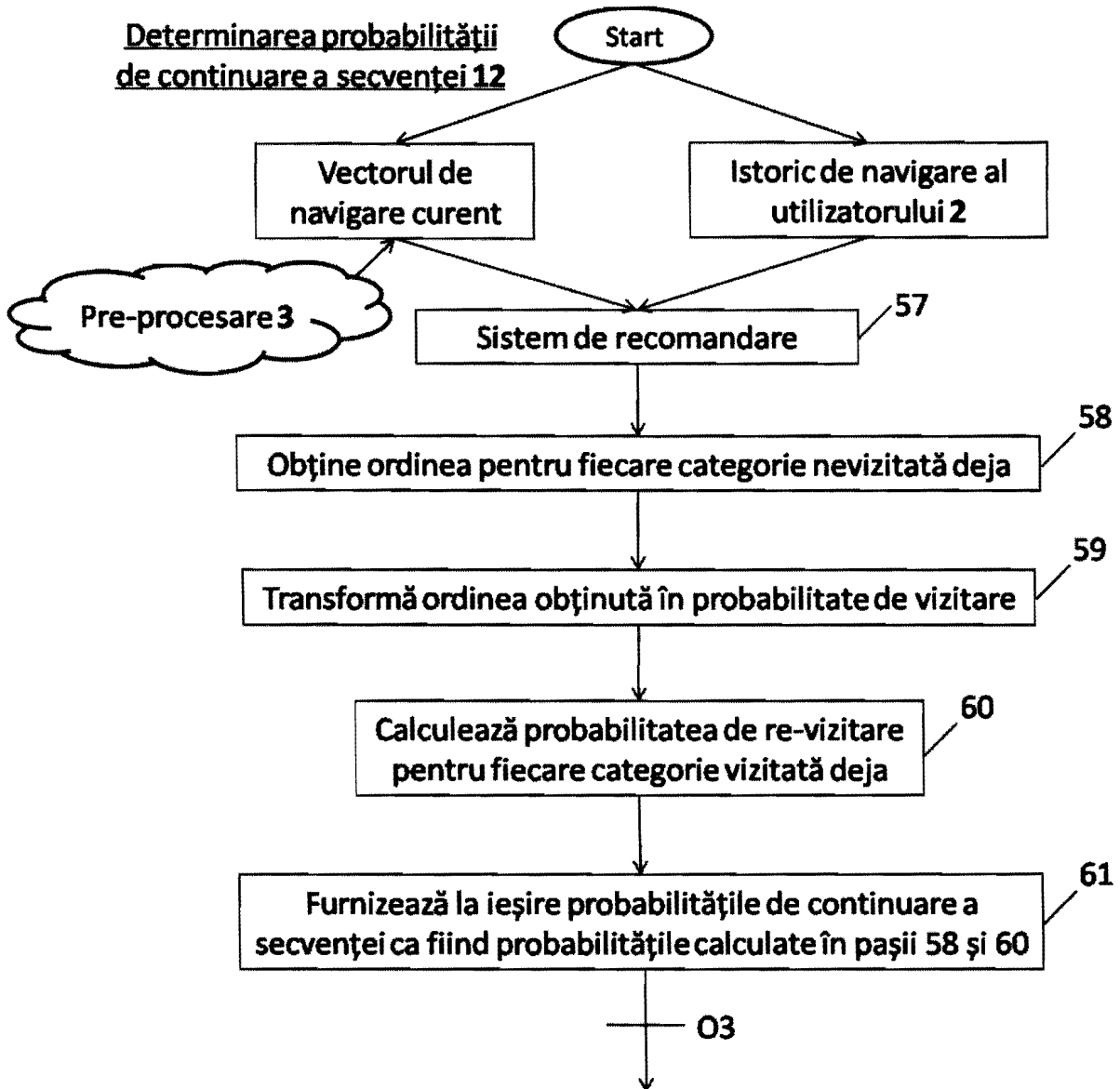


Fig. 8