

REFERAT

Privind teza de doctorat intitulată **CONTRIBUȚII LA DEZVOLTAREA PORTALURILOR PRIN INTERMEDIUL SISTEMELOR DE RECOMANDARE A CONȚINUTULUI** elaborata de domnul Daniel MICAN

Subsemnatul Ion IVAN profesor doctor la Departamentul de Informatică și Cibernetică Economică din Academia de Studii Economice, numit în comisia de susținere publică a tezei de doctorat prin Decizia Directorului CSUD de la Universitatea Babeș-Bolyai din Cluj-Napoca cu nr. 37010 din 19 decembrie 2012, am citi cu deosebită atenție teza de doctorat **CONTRIBUȚII LA DEZVOLTAREA PORTALURILOR PRIN INTERMEDIUL SISTEMELOR DE RECOMANDARE A CONȚINUTULUI** elaborată de domnul Daniel MICAN și în continuare prezint concluziile la care am ajuns.

Tema abordată în această teză de doctorat prezintă o importanță deosebită pentru cercetarea din domeniul informaticii economice, prin faptul că societatea informațională dezvoltată presupune generalizarea interacțiunilor dintre organizații, soluționarea de probleme complexe de către persoane fizice, precum și asigurarea unui management eficient a conținutului digital. În contextul creșterii eficienței sistemelor informatice complexe, toate aspectele de optimizare vizează structurarea informației, asigurarea unui nivel cât mai ridicat de securitate și creșterea numărului și diversității celor care interacționează cu aplicațiile informatice. Aspectele cele mai importante sunt legate de calitatea conținutului digital, calitate dată în procesul de creare a acestui nou tip de conținut, dar mai ales prin asigurarea acurateții proceselor de actualizare a conținutului digital. Autorul acestei teze a studiat în profunzime problematica dezvoltării portalurilor și experiența proprie în această zonă îl recomandă ca un valoros specialist în dezvoltarea de procese de proiectare, dezvoltare, implementare, utilizare, mentenanță și reinginerie de portaluri. Acest lucru este susținut de bibliografia inclusă în teză, caracterizată prin modul exigent cu care au fost selectate titlurile, actualitatea problematicilor abordate în cărți, în articole și în comunicările prezentate la conferințe științifice, publicate în volume, precum și site-urile de specialitate și forumurile care abordează problema conținutului digital construit prin sistemul recomandărilor.

Teza este o construcție bine documentată, în structura căreia autorul ei Daniel MICAN, a inclus capitole care îi permit o abordare gradată a problematicii. Astfel în prima parte, sunt incluse acele aspecte ce permit lui Daniel MICAN să prezinte realizările din domeniul cercetat. Documentarea sa aprofundată i-a permis o trecere în revistă a celor mai importante realizări din domeniul dezvoltării portalurilor folosind sistemul de recomandare a conținutului, în țară noastră dar și din străinătate. Citările sunt alese cu grijă și analizele sunt complete, autorul tezei Daniel MICAN venind cu o serie de considerații originale, identificând acele aspecte pe care trebuie să le dezvolte prin cercetări proprii.

Partea a doua a tezei, conține soluțiile originale obținute de Daniel MICAN în stagiul doctoral. Autorul prezentei teze se prezintă ca un cercetător format care știe cu maturitate să dezvolte un proces de cercetare, pornind de la formularea problemei, realizarea unei documentări sistematice, căutarea de noi soluții folosind tehnici și metode moderne din domeniul IT, verificarea soluțiilor pe cale experimentală și analiza acestor soluții, pentru a utiliza numai acele soluții care se dovedesc atât originale, cât și eficiente.

Partea a treia a tezei include aspectele de diseminare, Daniel MICAN dovedind un interes deosebit pentru a face cunoscute rezultatele sale atât în lumea specialiștilor informaticieni cât și la nivelul organizațiilor, prin implementări efective la nivelul unor portaluri a sistemului de recomandări pentru obținerea unui conținut digital de înaltă calitate.

Teza este structurată pe mai multe capitole și în continuare voi analiza fiecare capitol în parte.

În **INTRODUCERE** este definit scopul întregii teze și sunt precizate care sunt obiectivele stagiului de cercetare. Sunt analizate portalurile ca tipologii, ca structuri și sunt detaliate modalități de proiectare și realizare a acestora în contextul creșterii accelerate a complexității interacțiunilor cu grupuri neomogene de utilizatori. Este făcută o analiză detaliată a problematicii definirii și realizării conținutului digital, cu luarea în considerare a sistemelor de recomandare, singurele care au capacitatea de a asigura acel nivel de calitate care să dea operaționalitate conținutului digital, știut fiind faptul că acum pe baza conținutului digital se efectuează tranzacții, se alocă resurse și se iau decizii majore. Sunt analizate modalități de construire a sistemelor de recomandare bazate pe reguli de asociere și autorul propune un sistem propriu de recomandare folosind informații legate de dinamica comportamentului fiecărei persoane care participă la crearea de conținut digital într-o rețea. Sistemul este cu un grad de generalitate foarte ridicat ceea ce îl recomandă pentru orice structură de portal, dacă se urmărește crearea de submulțimi de utilizatori cu grad ridicat de omogenitate comportamentală. Tot în introducere este prezentată teza, fiecărui capitol fiindu-i alocată o importanță prezentare, ceea ce permite crearea unei imagini complete față de întregul demers științific al autorului.

În capitolul **Portaluri, CMS-uri, Web 2.0 și sisteme de recomandare** este prezentat conceptul de portal, autorul făcând o incursiune în evoluția acestui concept strict legată de progresele în timp ale tehnologiilor IT&C, cu prezentarea avantajelor pe care le aduce implementarea lor la nivelul unor organizații, atât în ceea ce privește capacitatea persoanelor de a se informa, cât mai ales modalitățile de realizare a unor procese complexe de alocare de resurse, de efectuare de tranzacții, de instruire și de realizare de interacțiuni în timp real. Este efectuată o clasificare a portalurilor și autorul tezei, Daniel MICAN dezvoltă o analiză detaliată legată de stabilirea funcțiilor unui portal. Experiența sa arată cu claritate că stăpânește până la nivelul detaliilor acele elemente legate de tehnicile de realizare și implementare a componentelor care realizează fiecare funcție în parte, fie că este vorba de prezentare, de calcul, de preluare de mesaje cu validare și de lucru pe baze de date, preluând cele mai noi realizări în ceea ce privește schimburile de mesaje dintre aplicație și utilizatori. Pornind de la analiza teoriei portalurilor este prezentată o arhitectura de portal cu luarea în considerare a componentelor și funcționalităților dezvoltate de acestea, în contextul fluxurilor care urmăresc menținerea redundanței între limite de acceptabilitate. Datorită faptului că fiecare portal presupune existența unui conținut digital structurat, cu niveluri diferite de variabilitate și cu modalități de prezentare neomogene specifice aplicațiilor multimedia, în teză sunt prezentate în detaliu toate elementele ce definesc conținutul digital, împreună cu sisteme care asigură managementul acestui conținut (Content Management Systems). Daniel MICAN realizează o prezentare a sistemelor de management de conținut, cu prezentarea unor puncte de vedere proprii, pornind de la funcțiunile și arhitecturile sistemelor de management de conținut. Conținutul digital este profund influențat de apariția Internetului și de dezvoltarea tehnologiilor web, lucru reflectat în mod corespunzător în teză, întrucât se face distincție între date, informații și cunoștințe și autorul ține seama de procesele de transformare a datelor în informații și a informațiilor în cunoștințe, procese esențiale în dinamica portalurilor moderne, în care sunt implementate sisteme de recomandare pentru asigurarea calității conținutului. În contextul actual, pentru dezvoltarea de portaluri există tehnologii asupra cărora autorul se

oprește în teză și face o analiză aprofundată. Sunt luați în considerare factori care definesc generatorii de conținut personalizat, care reflectă niveluri de inteligență diferite, urmărindu-se ca prin definirea sistemelor de recomandare să se atenueze toate aceste diferențe care în final se concretizează prin diferențe de calitate la nivelul conținutului digital în ansamblul lui. Crearea de modalități automatizate de stocare și de analiza inteligentă a conținutului sunt premise realiste de dezvoltare a portalurilor cu structură liberă de conținut generată de persoane acceptate prin sisteme de recomandare care asigură acuratețea și consistența tuturor proceselor de compunere sau agregare a conținutului.

Analiza funcționalităților are la baza seturi de date obținute, prelucrate și prezentate sugestiv de autor, referitoare la nivelul cunoștințelor în domeniul IT (figura 24), frecvența de utilizare al unui WCMS (figura 25), folosirea limbajelor de programare (figura 26), cele mai importante funcționalități în crearea conținutului (figura 27), cele mai importante funcționalități în managementul conținutului (figura 28), cele mai importante funcționalități în publicarea conținutului (figura 29), factorii care influențează în decizia de alegere unui WCMS (figura 30), preferințele pentru WCMS (figura 31) și modul de folosire a WCMS în realizarea unei comunități online colaborative (figura 32), i-a permis acestuia construirea unei arhitecturi eficiente pentru definirea unui sistem de recomandări care să conducă la asigurarea nivelului ridicat al calității conținutului în portaluri de complexitate ridicată. Amploarea analizei a condus la includerea în arhitectură a componentelor ce preiau fluxuri specifice interacțiunii cu utilizatorii dar și a fluxurilor interne ale portalurilor, realizând în acest fel o tratare completă în condițiile existenței de mecanisme de filtrare puternice a elementelor neconforme ce intră pentru a se constitui în componente de conținut digital.

În capitolul **Sistemul de recomandare WRS bazat pe reguli de asociere** autorul are în vedere descrierea și contextul problemei de construire a unui sistem de recomandare, formalizarea modelului propus, realizarea unei arhitecturi generale, prezentată în figura 33, proiectarea modulului de colectare, filtrare și extragere a regulilor de asociere și a modulului de recomandare. Sunt tratate aspecte legate de implementarea WRS-Wise Recommender System, de identificare a roboților și a motoarelor de căutare, de identificare și evidențiere a utilizatorilor și a paginilor. Sunt prezentate modalități de construire a sesiunilor de navigare, de extragere a regulilor de asociere și de preparare și de livrare a listei cu recomandări.

În acest capitol este evaluată scalabilitatea WRS, extragerea regulilor de asociere în WRS și compararea cu alte abordări.

Sunt analizate Sesiunile de navigare pe baza cărora se vor extrage regulile de asociere și Valorile pentru suport și încredere calculate prin WRS.

Dacă se realizează un experiment pe un set mai mare de date se folosesc la intrare surse de date cu sesiunile de navigare ale utilizatorilor de pe diferite site-uri de mari dimensiuni în care utilizatorii definesc conținut digital propriu. Dacă se realizează un export în timp real al sesiunilor de navigare și compararea rezultatelor cu cele furnizate printr-un WRS se accesează interfețele realizate prin aplicația propusă pentru export în formatul text sau cel al utilizării virgulei ca separator de entități.

Autorul propune (subcapitolul 3.2 **FORMALIZAREA MODELULUI PROPUȘ**), un model cu grad ridicat de generalizare pentru toate tipologiile de portaluri, în care definește mulțimea P a paginilor și extrage subseturi de pagini. Este definită regula de asociere a paginilor cu conținut total diferit și propune o modalitate interesantă de a calcula *suportul* (support) și *încrederea* (confidence), unde *suportul* (pi), este egal cu numărul total de apariții ale paginii raportat la numărul total de tranzacții, iar încrederea unei reguli măsoară procentul în care un articol depinde de altul. Metricile propuse de autor pe parcursul întregii teze sunt caracterizate prin consistență, iar analiza dimensională evidențiază caracterul reprezentativ ale tuturor

indicatorilor. Trebuie remarcat faptul ca indicatorii au la baza date ce se culeg automat pe măsură ce se derulează procesele de creare și actualizare a conținutului digital.

Capitolul **Sistemul de recomandare social WSNRS bazat pe scoruri de încredere** abordează aspecte de detaliu legate de rezultatele cercetărilor întreprinse de autor în ceea ce privește definirea unui sistem de recomandare care să fie bazat pe analiza automata a comportamentului fiecărei persoane, comportament dat prin structurile de conținut digital, prin modul în care interacționează în cadrul portalului ca frecvență, ca lungime de text, ca diversitate a tagurilor. Sistemele WSNRS - Wise Social Network Recommender System, fac obiectul unei abordări sistematice în teză și sunt dezvoltate de autor, pentru a atinge obiectivele propuse. Sunt analizate interacțiunile, receptivitatea la noi segmente de conținut digital și la modalități directe dar și la modalități indirecte de realizare a proceselor de recomandare. Sunt analizate modalități prin care se reflecta reacțiile care se propagă prin introducerea pe portal a noi elemente de conținut digital, în contextul stabilității acestui conținut. În cazul structurilor de socializare, în care interacțiunile sunt dinamice, numeroase și mai ales cu caracter aleator, dificil de gestionat în timp real, problema definirii de profiluri pentru generatorii de conținut crește în complexitate, ceea ce impune concentrarea de informație cu etichetarea de grupuri cu grad de omogenitate stabilit ca limită de acceptabilitate. Formalizările introduse de autor în capitolele precedente sunt verificate și se demonstrează că toate construcțiile teoretice își dovedesc valabilitatea. Aspectele experimentale evidențiază caracterul realist, riguros fundamentat pentru toate soluțiile date de autor.

Este definită și formalizată relația utilizator-utilizator, prin definirea de perechi, în care sunt luate în considerare tipul legăturilor, numărul de surse favorite, numărul notărilor pozitive, numărul de recomandări al utilizatorului, numărul de implicări prin analiza de comentarii, numărul de click-uri, descrierea interacțiunii prin amprenta de timp și scorul de încredere acordat de un utilizator celui alt utilizator.

Arhitectura descrisă în figura 46 și analizată în detaliu reprezintă un rezultat valoros întrucât ia în considerare toate ipotezele avansate în procesul de cercetare pentru a asigura un sistem de recomandare general valabil pentru o mare varietate de portaluri. Formalizările legate de interacțiunile utilizator-utilizator și utilizator-conținut vin să nuanțeze dinamica proceselor de management de conținut digital, întrucât aspectele de recomandare se dezvoltă dinamic și sunt determinate de grupurile care interacționează într-o rețea închegată prin structura portalului.

Problematica din capitolul **Concluzii generale și perspective ale cercetării** are menirea de a concentra cele mai importante aspecte incluse în teză. Autorul tezei, Daniel MICAN face un rezumat al soluțiilor originale detaliate în teză și prezintă o serie de direcții de cercetare viitoare, dintre care am remarcat:

- crearea unui model de sistem de recomandare ce poate fi folosit pentru furnizarea de conținut, sugerarea de utilizatori și livrarea de material publicitar personalizat în cadrul portalurilor web;
- datele colectate vor fi analizate cu ajutorul tehnicilor de data mining în vederea extragerii unor modele pe baza cărora să se poată face predicții și recomandări. Sistemul pe care dorim să îl propunem este o abordare hibridă ce dorește să folosească tagurile pentru crearea poștelor contextuale ale utilizatorilor împreună cu tehnici de filtrare colaborativă;
- studierea modului în care încrederea este transferată utilizatori, respectiv între utilizatori la resurse;
- îmbunătățirea modelului prin integrarea unui sistem de recomandare cu ajutorul analizei tagurilor;

- identificarea experților și a conținutului generat, bazat pe diferite categorii de interes
- realizarea unei agregări hibride prin preluarea elementelor din sistemul, bazat pe extragerea regulilor de asociere din cadrul sesiunilor de navigare.
- orientarea cercetărilor către propunerea unui sistem hibrid care să soluționeze problemele care afectează calitatea conținutului, apărute în sistemele de recomandare clasice.

Bibliografia care stă la baza întregului demers științific al lui Daniel MICAN este formată din titluri de cărți reprezentative, include articole publicate în reviste indexate în baze internaționale importante, precum și lucrări realizate de autor și în colaborare. Cea mai mare parte a lucrărilor incluse în bibliografie sunt publicații din ultimii cinci ani, lucru care evidențiază preocuparea autorului de a fi la zi cu cele mai noi realizări din domeniu.

Teza include și o serie de anexe. Anexa A este dedicată analizei comparate între WCMS-uri și funcționalități unde autorul pentru opt produse realizează descrieri în raport cu cerințele serverelor, costurilor, bazelor de date, sistemelor de operare, limbaje de programare, verificare e-mail, autentificare, management server, compatibilitate SSL, comunicare cu dezvoltatorii, încărcare de fișiere, precum și o serie de indicatori măsurați pentru a caracteriza flexibilitatea produselor supuse analizei. Efortul de sistematizare și analiza inclusă în teză bazată pe aceste date au permis obținerea unor concluzii foarte importante pentru ceea ce se dorește să se obțină prin sistemele de recomandare în zona asigurării calității conținutului digital, în contextul în care acesta stă la baza derulării de activități, se alocă efectiv resurse și pe baza căruia se iau decizii vitale pentru persoane și pentru organizații.

Anexa B este destinată prezentării chestionarului de culegere a datelor privind analiza WCMS-urilor și a funcționalităților, Daniel MICAN incluzând aspecte de definire a respondenților, a calificării pe niveluri a modului în care se realizează interacțiunile cu produsele de tip WCMS. Apreciez că Daniel MICAN a realizat o structură completă a chestionarului, astfel încât să se obțină acea formă care este ușor de urmărit, ușor de completat și mai ales, care să creeze o imagine completă a modului în care concret sunt utilizate produsele și mai ales cum acestea contribuie la asigurarea calității conținutului obținut prin sistemul de recomandare, știut fiind faptul că într-un portal complex există foarte multe persoane care contribuie la actualizarea conținutului și mai ales faptul că aceste persoane sunt neomogene că experiența și nivel de cunoștințe și comportament dar mai ales structura grupurilor care actualizează conținutul digital este foarte dinamică datorită fluctuațiilor de personal din organizații.

Anexa C conține regulile de asociere generate cu ajutorul programului weka.

În raport cu criteriile care stau la baza evaluării unei teze de doctorat, rezultă următoarele:

- Criteriul privind structurarea tezei este integral îndeplinit, pentru că lucrarea conține o introducere în care sunt stabilite obiectivele cercetării, sunt prezentate aspectele pe care autorul le are în vedere pe durata întregului său demers științific; teza include capitole în care sunt evidențiate concepte, tehnici și metode de creare a conținutului digital și modalități de asigurare a calității în procesul de mentenanță atât a portalurilor, cât și a conținutului digital în condițiile de dinamică a tehnologiilor informaticii dar și a fluctuațiilor de personal; autorul a studiat o bogată literatură de specialitate încât cunoaște realizările cercetătorilor din țară și din străinătate; teza include în capitolele sale prezentarea soluțiilor originale obținute pe parcursul stagiului doctoral, autorul dând dovadă de un nivel de exigență ridicat, ceea ce i-a permis selectarea și includerea în teză numai a soluțiilor originale eficiente și reprezentative; în teză sunt oferite detalii asupra modului în care s-a efectuat diseminarea rezultatelor și sunt prezentate direcții de cercetare pe care autorul tezei, Daniel MICAN își propune să le urmeze în viitor; lucrarea are o bibliografie bogată și două anexe; autorul stăpânește tehnicile realizării unei lucrări de anvergură, prin faptul

că gestionează foarte bine modul în care a efectuat citările, definirea și referirea de figuri, diagrame și tabele și asigură consistență modelelor introduse de el prin modul în care a construit sistemul de identificare prin variabile a factorilor de influență a fenomenelor studiate;

- Criteriul privind soluțiile originale incluse în teză este **integral îndeplinit**; autorul acestei teze, Daniel MICAN, a obținut pe parcursul stagiului de cercetare soluții originale privind:
 - realizarea unei metode de examinat sistemele de personalizare bazate pe analiza fișierelor log
 - propunerea de metode proactive concretizate printr-un sistem de colectare exclusivă a datelor folositoare procesului de web usage mining
 - elaborarea unei tehnici cu ajutorul căreia se identifica în mod unic fiecare utilizator ținând cont atât de adresa IP a serverului proxy cât și a stației de lucru
 - realizarea unei modalități de investigare a problematicii extragerii regulilor de asociere
 - obținerea unei abordări care permite extragerea și stocarea în mod tranzacțional și incremental a tuturor regulilor de asociere
 - construirea unei arhitecturi originale care permite stocarea relațiilor stabilite între conținutul accesat în mod ocazional fără a afecta scalabilitatea sistemului
 - dezvoltarea, implementarea și testarea unei abordare care permite identificarea și recomandarea resurselor de valoare în cadrul structurilor sociale
 - elaborarea unui algoritm de calculare a scorurilor de încredere între utilizatori ținând cont de interacțiunile pe care aceștia le realizează atât între ei cât și cu resursele publicate
 - propunerea unei metode proprii de deducere a relațiilor implicite de tip adept între utilizatori unei rețele sociale folosind scorurile de încredere calculate
 - analiza beneficiilor aduse de abordarea proprie față de abordările clasice de recomandare colaborativă și cea bazată pe conținut datorită faptului că permite recomandarea resurselor ținând cont numai de structura socială și relațiile utilizatorilor
 - a fost dezvoltat și implementat un sistem de etichetare colaborativă cu ajutorul tagurilor
 - a fost studiată, pe baza datelor extrase, oportunitatea dezvoltării unei extensii contextuale pentru WSNRS
 - a fost dezvoltată tehnica de recomandare de resurse etichetate cu ajutorul tagurilor
 - au fost identificate și folosite modalități de calcul a similarității existente între taguri și s-a demonstrat acestea dacă sunt folosite, contribuie la creșterea eficienței sistemelor de recomandare.
- Criteriul privind diseminarea rezultatelor cercetării este integral îndeplinit de către autorul tezei Daniel MICAN, întrucât a publicat articole, a fost prezent la conferințe științifice de specialitate, cei prezenți apreciind soluțiile originale incluse în comunicările sale; toate aceste afirmații sunt sprijinite de titlurile materialelor elaborate și publicate de autor sau în colaborare, după cum urmează:
 - Web 2.0 as a New Vision of Web-Based Applications (2009)
 - Web 2.0 and Collaborative Tagging (2010)
 - Optimized Advertising Content Delivery (2008)
 - Building a Social Recommender System by Harvesting Social Relationships and Trust Scores between Users (2012)

- Association-rules-based recommender system for personalization in adaptive web-based applications (2010)
- Content Management System for Academic Management, InfoBusiness (2006)
- Preprocessing and Content / Navigational Pages Identification as Premises for an Extended Web Usage Mining Model Development (2009)
- Web Content Management Systems, a Collaborative Environment in the Information Society (2009).

Diversitatea manifestărilor științifice la care a fost prezent, complexitatea lucrărilor prezentate, îl definesc pe Daniel MICAN ca pe un cercetător format, cu un nivel de maturitate ridicat, exigent, care are toate elementele necesare dezvoltării de proiecte de cercetare importante pentru că el știe să definească o problemă, știe să o descompună în subprobleme și are capacitatea de a forma o echipă performantă și de a coordona cu succes o astfel de echipă.

Toate acestea arată că lucrarea **CONTRIBUȚII LA DEZVOLTAREA PORTALURILOR PRIN INTERMEDIUL SISTEMELOR DE RECOMANDARE A CONȚINUTULUI** elaborată de domnul Daniel MICAN este o construcție unitară, rezultatul unei activități de cercetare într-un domeniu nou și autorul a reușit să urmărească pas de pas obiectivele stabilite și să soluționeze creativ problemele, verificând de fiecare dată în mod exigent, calitatea, eficiența și generalitatea soluțiilor sale.

Analizând lucrarea **CONTRIBUȚII LA DEZVOLTAREA PORTALURILOR PRIN INTERMEDIUL SISTEMELOR DE RECOMANDARE A CONȚINUTULUI** elaborată de domnul Daniel MICAN, am ajuns la următoarele concluzii: are un nivel științific deosebit, este bine structurată, include numeroase soluții originale, autorul a depus eforturi speciale în a asigura un proces de diseminare ritmic și susținut. Toate acestea mă îndreptățesc să fiu de acord ca teza să fie susținută în ședință publică și autorului ei Daniel MICAN să-i fie atribuit titlul de doctor în științe, în domeniul fundamental economic, specializarea cibernetică și statistică economică, cu calificativul **FOARTE BINE**.

6 ianuarie 2013

Prof.dr Ion IVAN