



Universitatea Tehnică a Moldovei

**PLATFORMĂ WEB DE PREZICERE A EVOLUȚIEI
ACȚIUNILOR COMPANIILOR MARI**

**WEB PLATFORM FOR PREDICTING THE PROGRESS
OF THE ACTIONS LARGE COMPANIES**

Student:

Dan Caraganciu

Conducător:

**Galina Marusic
dr., conf.univ.**

Chișinău, 2020

Adnotare

Structura lucrării: Teza de master este alcătuită din introducere, 3 capitole care descriu cercetarea temei alese, concluzii și bibliografie, 70 de pagini, de text și 45 de figuri. În capitolul 1 este analizat domeniul de cercetare modalitățile de procesare a documentelor electronice, principalele caracteristici ale sistemelor de descoperire a resurselor. În capitolul 2 sunt descrise principalele metode și tehnici de prelucrare a datelor, de clasterizare și modalități de diseminare a rezultatelor. Capitolul 3 reprezintă partea practică a tezei, unde sunt descrise principalele modalități de creare a aplicației. Tot aici este prezentată descrierea aplicației elaborate. În concluzii sunt prezentate principalele idei ce rezidă din tema cercetată.

Cuvinte-cheie: tehnologii informaționale, regăsirea informației, structurarea informației, prelucrarea statistică a datelor, metode de grupare a datelor, API.

Scopul lucrării: elaborarea unei platforme Web de prezicere a evoluției acțiunilor companiilor mari.

Metodele aplicate la realizarea cercetării: În prezenta lucrare s-a plecat de la premisa utilizării metodelor pur computaționale în regăsirea informațiilor și structurarea informației pentru prezicerea evoluției acțiunilor companiilor mari.

Rezultatele obținute: Sistemul realizat este capabil să determine corect sentimentul fiecărui tweet și astfel permite prezicerea evoluției acțiunilor în timpul apropiat.

Valoarea aplicativă a tezei: rezultatele obținute pot fi utilizate de către acționari pentru a determina momentul potrivit de a cumpăra sau vinde acțiunile de la o companie sau alta.

Abstract

Thesis structure: The masters thesis includes the introduction, 3 chapters that describe the research of the chosen topic, conclusions, bibliography, 70 pages of text and 45 figures. Chapter 1 analyzes the field of research, the ways of processing electronic documents, the main features of resource discovery systems. Chapter 2 describes the main methods and techniques of data processing, clustering and ways of disseminating the results. Chapter 3 represents the practical part of the thesis, where the main ways to create the application are described. Also here is presented the description of the elaboration application. In the conclusions are presented the main ideas that reside from the researched topic.

Cuvinte-cheie: information technologies, information retrieval, information structuring, statistical data processing, data grouping methods, API.

Thesis scope: development of a Web Platform for predicting the evolution of stocks for big companies.

Applied methods during investigation process In this paper we started from the premise of using purely computational methods in retrieving information and structuring information to predict the evolution of the actions of large companies.

Results obtained: The system that was developed is capable of determining precisely the tweet sentiment and predicting the evolution on stocks in the near future.

The applicative value of the thesis: the results obtained can be used by shareholders to determine the right time to buy or sell shares from one company or another.

Cuprins

INTRODUCERE	8
1 ANALIZA METODELOR STATISTICE COMPUTAȚIONALE PRIVIND PRELUCRAREA DATELOR	9
1.1 Procesarea automată a documentelor de tip text	9
1.2 Text mining	14
1.3 Analiza și modelarea a orientării multi-agent	20
1.4 Sisteme pentru descoperirea de resurse.....	23
1.5 Formularea sarcinii și determinarea cerințelor.....	25
2 TEHNICI ȘI METODOLOGII DE EXPLORARE A DATELOR	28
2.1 Principii ale observării statistice	28
2.2 Tehnologii moderne de explorare și exploatare a datelor	31
2.3 Algoritmi de căutare cu revenire	39
2.4 Algoritmi de clasificare a documentelor	46
2.5 Metode ierarhice.....	51
2.6 Metode bazate pe ordinea cuvintelor.....	54
3 PROIECTAREA PLATFORMEI PRIVIND PREZICEREA ACȚIUNILOR	56
3.1 Alegerea strategii de proiectare	56
3.2 Schema conceptuală a aplicației.....	60
3.3 Crearea bazei de date a aplicației	64
3.4 Descrierea modulelor aplicației.....	68
3.5 Descrierea aplicației	70
CONCLUZII	73
BIBLIOGRAFIE	74
ANEXA A	75

INTRODUCERE

Domeniul gestiunii informațiilor abordează problema organizării, stocării și regăsirii în timp util a datelor de care dispunem despre un anumit subiect. Neîndoielnic că fiecare dintre noi se confruntă continuu cu această problemă. Agenda personală, cartea de telefoane, sunt mijloace simple de organizare, păstrare și regăsire a datelor de care avem nevoie la un moment dat.

Chiar dacă proiectanții interfețelor și a conținutului web-ului includ comportamentul utilizator și câteva deprinderi în unele motoare de căutare, utilizatorii pot avea dificultăți în efectuarea de căutări pe web deoarece aceștia nu sunt capabili să formuleze interogările corecte care să reducă numărul de rezultate obținute și să crească calitatea acestora. De obicei interfața utilizator este dificil de utilizat, funcțiile de organizare ierarhică sunt aplicate static și informațiile de pe web sunt foarte rar structurate și organizate pentru o recunoaștere rapidă.

Pentru a rezolva problemele prezentate mai sus acest domeniu de cercetare a crescut continuu în direcții cum ar fi algoritmi, strategiile și arhitecturile. Sute de idei au fost testate, unele fiind implementate iar altele existând doar ca și prototipuri. Unele dintre aceste idei sunt: organizarea documentelor în categorii predefinite, îmbunătățirea modului de prezentare a rezultatelor, monitorizarea unei pagini specifice, ajutarea utilizatorului în formularea interogării corecte și dezvoltarea de programe pentru filtrarea rezultatelor căutării.

O dată cu dezvoltarea rapidă a tehnologiilor informaționale a apărut necesitatea dezvoltării sistemelor informaționale, pentru a programa sisteme soft care ar putea stoca cantități enorme de informații și ar putea reda prin comenzi simple și ușor de asimilat date exacte despre conținutul bazei de date, precum și prelucrarea unor altor tipuri de date nefiind obligatoriu să prelucreze date sub formă de tabele dar pot fi date grafice sau date de avertizare sau de alt tip din domeniul informaticii, care sigur este domeniul cel mai important la ziua de azi precum și cel mai cercetat care datorită posibilități vaste pe care le poate îndeplini un soft combinat cu hard. Astfel viitorul zilei de mâine al oricărui domeniu nu poate fi conceput fără folosirea unui soft al domeniului respectiv.

În prezenta lucrare s-a plecat de la premisa utilizării metodelor pur computaționale în regăsirea informațiilor și structurarea informației pentru prezicere a evoluției acțiunilor companiilor mari. Cu toate că în unele cazuri s-a încercat adăugarea unei anumite „cantități de informație semantică” în diverșii algoritmi utilizați, abordarea din punct de vedere semantic depășește cadrul prezentei teze.

Reiesind din cele menționate s-a propus de elaborat un sistem pentru colectarea și procesarea comentariilor de pe Twitter în scopul determinării ratei comentariilor pozitive/negative pentru a efectua o statistică și în același timp pentru a încerca să prezic evoluția stocurilor în viitorul apropiat pe baza informației colectate.

Bibliografie

- 1 Guandong Xu, Zanchun Yhang, Lin Li, Web Mining and Social Networking Techniques and Applications, USA:Springer, 2011
- 2 J. Palau, M. Montaner, B. Lopez and J.L. de la Rosa. Collaboration analysis n the recommender system using social networks. In CIA, pages 137-151,2004
- 3 J.M. Kleinberg. Authoritative sources in a hyperlinked environment. In Proc. Of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '08), pages 668-677, 2008
- 4 Zaiane O., Han J.: WebML: Querying the World Wide Web for resources and knowledge. In: Workshop on Web Information and Data Management WIDM98, Bethesda, 1998, 9-12.
- 5 Yang, Q., Zhang, H.H. and Li, I.T, 2001. Mining Web logs for prediction models in www caching and prefetching, Seventh ACM SIGKDD International Conference on Knowledge
- 6 Voinea, Gheorghe Mecanisme și tehnici valutare și financiare internaționale, Editura Sedcom Libris, Iași, 2014.
- 7 Roșea I., Proiectarea sistemelor informatice financiar contabile, Editura Didactică și Pedagogică, București, 2005.
- 8 Andreea Saseanu Incertitudinea Strategica și Viitorul Dialogului On-line cu Consumatorul, Editura ASE, Bucuresti, anul 2015.
- 9 Dunham, M.H.: Data Mining: Introductory and Advanced Topics. Prentice Hall, Pearson Education Inc. (2013)
- 10 Agrawal, R., Srikant, R.: Mining sequential patterns, *International Conference on Data Engineering(ICDE'05)*, Taipei, Taiwan, pp. 3-14, martie 2005.
- 11 Schechter, S. Krishnan, M. Smith, M.D. :*Usingpathprofiles topredict http request, Seventh International World Wide Web Conference*, Brisbane, Australia, pp. 457-467, Aprilie, 2008.
- 12 Nong, Y.: The handbook of Data Mining, Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey, 2013.
- 13 Vercellis, C. : *Business Intelligence: Data Mining and Optimization for Decision Making*,UK: John Wiley & Sons, 2009
- 14 Jiawei Han, Micheline Kamber: *Data Mining Concepts and Techniques* USA: Elsevier, 2016
- 15 Stephanie Mattingly ,*Clickstream Analysis:Both a Business and an Aid in Advertising*