

Aplicație bazată pe microservicii pentru identificarea persoanelor cu interese similare

Radu-Valentin Cornea
UNIVERSITATEA TEHNICĂ
„Gheorghe Asachi” din IAȘI
FACULTATEA DE AUTOMATICĂ
ȘI CALCULATOARE
Iași, România
radu-valentin.cornea@student.tuiasi.ro
Coordonator științific:
Ș.I. dr. inf. Tiberius Dumitriu

Abstract—În această lucrare se propune prezentarea unor metode de recomandare a anumitor persoane, în funcție de preferințele pe care le au în comun cu ceilalți utilizatori, folosind metode din sfera inteligenței artificiale. Accentul cade pe grupuri de interes din sfera educațională, dar poate fi extins la orice alte domenii.

Studiul redă câteva aspecte generale ale aplicațiilor bazate pe microservicii, modalități de securizare a aplicației, tehnologiile folosite, motivația realizării unei aplicații diferite de cele existente, arhitectura realizată și rezultatele obținute. Experimentele efectuate presupun aplicarea unor algoritmi de inteligență artificială, precum k-nearest neighbors (KNN) pentru a i se recomanda unui nou utilizator, pe baza preferințelor sale, un grup de persoane (cunoscute sau nu) ce au interese similare. Pentru aceasta se utilizează valori diferite ale parametrului K, precum și mai multe metrici (Jaccard, euclidiană, cosine etc.) cu scopul de a determina cea mai potrivită recomandare. Rezultatele obținute sunt prezentate sub formă de tabele și grafice care evidențiază avantajele acestei abordări.

Pe baza experiențelor anterioare obținute în urma recomandărilor, unui utilizator nou i se poate asocia rapid un grup de persoane cu interese similare.

Cuvinte cheie—Aplicații software, Aplicații educaționale, Inteligență artificială

I. INTRODUCERE

Proiectul constă într-o aplicație pentru interconectarea oamenilor, în funcție de preferințele lor, utilizând diverși algoritmi. Unii dintre algoritmi utilizați sunt mai simpli, iar alții folosesc inteligență artificială. Se dorește și recomandarea utilizatorilor în funcție de alte filtre precum distanța geografică, dar momentan recomandările se fac doar în funcție de preferințe. Scopul proiectului este de a uni oamenii mai ușor în scop educațional în funcție de preferințele legate de tehnologii sau concepte teoretice, însă proiectul ar putea fi folosit și în alte arii decât cele educaționale.

Totuși, există o mulțime de studii pe tema recomandărilor, unii dintre oameni gândindu-se deja la sisteme de recomandări bazate pe inteligență artificială psihologia educației, tocmai pentru a recomanda studenților resurse de studiu, în funcție de cum interacționează cu imaginile de pe site-ul respectivei instituții și cu elementele video din acestea. Studenții au fost clasificați în: studenți activi, studenți cu potențial și studenți inactivi. Toate aceste trei grupuri au fost împărțite pe baza modului în care interacționau cu paginile respective, urmând ca fiecare dintre ei să primească resurse educative potrivite [1]. Există foarte multe domenii de aplicare pentru algoritmi

de recomandare, și pot fi folosiți nu doar în scopuri educative, ci și divertisment, precum filmele [2].

În urma unui studiu de piață, s-a constatat că, în ciuda faptului că există aplicații care încearcă să recomande oameni după anumite criterii, acestea nu realizează în totalitate obiectivul dorit. Multe dintre aplicațiile găsite pe Play Store includ opțiunea de alegere de preferințe (de exemplu, mâncare, hobby-uri, muzică), însă filtrele de utilizatori sunt inexistente, neputând primi utilizatori similari unei ținte, cel puțin din punctul de vedere al preferințelor. Singura aplicație care s-a constatat că ar face o parte dintre aceste funcționalități de recomandare este Meetup, dar acolo au loc recomandări de evenimente, nu de persoane. Panion ar fi fost un exemplu bun, dar în prezent nu mai funcționează publicului larg.

Pe baza acestui studiu de piață, s-a ajuns la concluzia că o astfel de aplicație pentru recomandarea persoanelor în funcție de preferințele lor ar fi necesară pieței. Obiectivul principal este acela de a găsi și de a filtra cât mai mulți utilizatori potriviți cu ținta în cauză, dar și de a avea un produs funcțional, sigur și securizat, care să fie ușor de folosit, plăcut și de înțeles de oricine.

O posibilitate de a rezolva această problemă este de a folosi microserviciile și inteligența artificială. Arhitecturile bazate pe microservicii oferă scalabilitate, rulare independentă, performanțe adiționale, avantajul de a fi ușor de menținut, costuri reduse și multe altele [3]. În timp ce inteligența artificială oferă posibilități multiple de modelare și antrenare a datelor după anumite seturi de date. Câțiva dintre acești algoritmi ce pot fi folosiți în acest sens sunt: KNN, SVM, SVD, Random Forest, filtrul colaborativ, filtrul bazat pe conținut, abordări hibride [4, 5].

Tocmai de aceea, în încercarea de a rezolva problema propusă, tehnologiile și conceptele teoretice folosite în cadrul aplicației pot fi clasificate astfel:

- Backend: Spring, Spring Security, JWTs, REST, Kotlin, MariaDB, MongoDB, criptare, decriptare;
- Frontend: Thymeleaf, JavaScript, CSS, HTML;
- AI: KNN, Python, sklearn, pandas;
- Aplicații software: IntelliJ IDEA, PyCharm, Visual Studio Code, Postman, MongoDB Compass, DBeaver.

II. IMPLEMENTARE (?)

A. Concept

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

B. Proiectarea aplicației

Pentru o bună funcționalitate, dar și dezvoltare mai facilă a aplicației, părțile de cod au fost despărțite pe module și clase, în unele situații fiind necesare șabloanele de proiectare. Unul dintre acestea este chiar șablonul de proiectare strategie. Acesta a fost folosit pentru a decide modul în care utilizatorii vor fi recomandați unei ținte. În prezent, sunt patru strategii de a decide acest aspect: căutări directe pe utilizatori utilizând paradigma funcțională, iar celelalte strategii bazându-se pe algoritmul KNN și câteva metrice ale acestuia.

Alte părți de cod au o granularitate fină, având propriul proiect și server, urmărindu-se în acest fel și decuplarea aplicației dezvoltate. Aplicația se folosește de modelul client-server, fiind bazată pe servicii/microservicii. Câteva dintre serviciile implementate până acum cuprind: IDM (Identity Management), Profile, Algorithms și Gateway-ul ce leagă serviciile. Alte exemple de servicii utile aplicației, dar neimplementate încă, sunt cele pentru identificarea locației, comunicarea între utilizatori, și pentru suport-ul utilizatorului (în cazul în care acesta are nevoie de ajutor, să poată depune cereri).

schema

C. Modalități de securizare a aplicației

Pentru a fi siguri că aplicația este una sigură în ceea ce privește datele utilizatorilor, s-au utilizat diverse mecanisme de protecție a datelor, precum JWT-urile, păstrate la nivelul clientului în format criptat (tocmai pentru a se evita schimbarea câmpurilor din interiorul lor), ele fiind mai întâi decriptate la nivelul server-ului, apoi se validează formatul și semnătura acestora, urmând abia apoi să se realizeze validările pe câmpurile lor (de exemplu, există un câmp expiry), autorizarea realizându-se abia la sfârșit, în funcție de autoritatea pe care o deține utilizatorul respectiv. Pentru a persista parolele, acestea sunt ținute în baza de date criptate cu ajutorul funcției BCrypt, iar în cazul în care un client se autentifică, la verificare, parola introdusă este criptată și ea și comparată cu valoarea din baza de date. S-ar fi putut folosi alte metode pentru autentificare și autorizare, precum OAuth2, SAML, OpenID, dar în scop demonstrativ s-au folosit doar JWT-urile.

D. Date și algoritmi (Structuri de date și algoritmi ??)

TODO

E. Algoritmi de inteligență artificială

Până în momentul de față, s-a realizat o comparație între câteva dintre metricele algoritmului KNN, acestea fiind Jaccard, euclidiană și cosine. Dar pe lângă aceste comparații, s-au întocmit și niște rapoarte care ar sugera diverse situații de utilizare mai potrivite acelor metrice.

F. Interfață grafică (1/2 – 1)

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template

measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. EXPERIMENTE ȘI REZULTATE

Cum sunt făcute

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Tabele și diagrame

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.

IV. CONCLUZII

Ce și de ce? Ce aduc în plus?

A. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

B. Figures and Tables

a) *Positioning Figures and Tables*: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^a Sample of a Table footnote. (Table footnote)

Fig. 1. Example of a figure caption. (figure caption)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

- [1] <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.767837/full>
- [2] <https://www.sciencedirect.com/science/article/pii/S0950584921000793?via%3Dihub>
- [3] <https://www.mdpi.com/2227-7390/10/7/1192>
- [4] <https://link.springer.com/article/10.1007/s40747-020-00212-w>
- [5] <https://link.springer.com/article/10.1007/s40747-020-00212-w>