

ობიექტ-ორიენტირებული ვებ-გვერდის კლასიფიცირების მაგალითი ძირითადი მახასიათებლების მიხედვით

გიორგი კაკაშვილი

საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

იქედან გამომდინარე, რომ მსოფლიო ქსელში წარმოდგენილი ვებ-გვერდები ჯეროვან დონეზე არ არის კლასიფიცირებული, რაც დადასტურებულია ჩვენ მიერ ჩატარებული კვლევებით, ვთვლით, რომ ვებ-გვერდებზე განთავსებულ ინფორმაციული ნაკადის მოწესრიგება უნდა მოხდეს მკაცრად განსაზღვრული კანონზომიერებით. ჩვენ მიერ შემოთავაზებულ მოდელში, რომლის მაგალითსაც აქ ვიხილავთ, ვებ-გვერდების კლასიფიცირების დროს ძირითადი აქცენტი მიმართულია ვებ-გვერდზე არსებული ინფორმაციის დანიშნულებას, ფასეულობასა და სენსიტიურობაზე. დამუშავებულია ვებ-გვერდის კლასიფიკაცია, როგორც პრობლემური ამოცანა და განსაზღვრულია ვებ-გვერდის კატეგორიები. აგებულია გადაწყვეტილების მატრიცა, რომლის მეშვეობითაც განისაზღვრება ვებ-გვერდის კლასი.

საკვანძო სიტყვები: ობიექტ-ორიენტირებული. ვებ-გვერდი. კლასიფიკაცია. ინფორმაციული ნაკადი. ინტერნეტსივრცე. სისტემური მიდგომა. გადაწყვეტილების მატრიცა.

1. შესავალი

ვებ-გვერდის კლასიფიცირებაში ვგულისხმობთ ვებ-გვერდების სიმრავლის ისეთ ჰარმონიულ სისტემას, რომელიც ქმნის სტრუქტურულად მოწესრიგებულ ერთიან საინფორმაციო სამყაროს ანუ კლასიფიცირებულ სამყაროს. ცხადია, ვებ-გვერდებზე განთავსებული ინფორმაციული ნაკადის მოწესრიგება უნდა მოხდეს მკაცრად განსაზღვრული კანონზომიერებით, რომლის დანიშნულებაც ინფორმაციის განთავსების, მოძიების, აღქმის, გაანალიზების და გამოყენების სიიარაღე, სიმარტივე, სრულყოფილი შინაარსი.

ჩვენ მიერ ჩატარებულმა კვლევებმა გვაჩვენა, რომ მსოფლიო-ქსელში წარმოდგენილი ვებ-გვერდები ჯეროვან დონეზე არ არის კლასიფიცირებული და სტრუქტურირებული, რაც ქმნის ერთგვარ ინფორმაციულ ქაოტს, განმეორებას და სიჭარბეს.

ვებ-გვერდების პროექტირების არაკლასიფიცირებული და არასისტემური მიდგომის მაგალითია ის, რომ ფუნქციურად სრულიად განსხვავებულ ორგანიზაციებს აქვს ინდენტური სტრუქტურის და დიზაინის ვებ-გვერდი. ცხადია, ასეთი საინფორმაციო ველის არსებობა ქმნის მრავალ პრობლემას, რომელთაგანაც მომხმარებლისათვის უმთავრესი დაბრკოლებაა ინფორმაციის მოძიების პროცესში ერთგვარი გაურკვევლობის შექმნა. ასეთ ვითარებაში მომხმარებელს უძნელდება სამიზნო ინფორმაციამდე მისაღწევი სვლაგეზის დადგენა და იგი ერთგვარ ლაბირინთის ტყვეობაში ექცევა. აღნიშნული სიტუაციის ანალიზის საფუძველზე დავადგინეთ, რომ საჭიროა დამუშავდეს ვებ-გვერდის კლასიფიცირების მოდელი, რომლის მეშვეობითაც მოხდება ინფორმაციული ველის მოწესრიგება მისი ძირითადი მახასიათებლების მიხედვით. ერთიანი ვებ-სივრცე, სადაც

კლასიფიცირებული იქნება ვებ-გვერდები ობიექტზე ორიენტირებული მახასიათებლების მიხედვით, მნიშვნელოვანწილად გადაწყვეტს აღნიშნულ პრობლემას.

ჩვენ მიერ შემოთავაზებულ მოდელში, რომლის მაგალითსაც აქ ვიხილავთ, ვებ-გვერდების კლასიფიცირების დროს ძირითადი ყურადღება გამახვილებულია ვებ-გვერდზე არსებული ინფორმაციის დანიშნულებაზე, ფასეულობასა და სენსიტიურობაზე. იქედან გამომდინარე, რომ ყველა ვებ-გვერდს ჰყავს თავისი მფლობელი და პასუხისმგებელი მასზე განთავსებული ინფორმაციის საიმედოობაზე, ვფიქრობთ, მათთვის საინტერესო იქნება ჩვენი შეთავაზება. შეთავაზება მდგომარეობს მასში, რომ მათ ვთავაზობთ კლასიფიცირებულ სივრცეში მოქცეულ მულტიფუნქციურ ვებ-გვერდს, რომლის საშუალებითაც შეძლებენ განსაზღვრონ თავიანთი ვებ-გვერდის რომელიმე კლასთან მიკუთვნება და ის მინიმალური სტანდარტები, რომელიც ლეგიტიმურს გახდის მათი ვებ-გვერდების არსებობას მისთვის ადექვატურ კლასში.

2. ძირითადი ნაწილი

განვიხილოთ ობიექტ-ორიენტირებული ვებ-გვერდის კლასიფიცირების მაგალითი. წარმოვადგინოთ ვებ-გვერდის კლასიფიკაცია, როგორც პრობლემური ამოცანა და განვსაზღვროთ ესა თუ ის ვებ-გვერდი რომელ კატეგორიას განეკუთვნება [1,3].

ამოცანას მივცეთ ფორმულიზებული სახე. შემოვიტანოთ აღნიშვნები:

$C = \{c_1, \dots, c_k\}$ - წინასწარ განსაზღვრული კატეგორიის კომპლექტი;

$D = \{d_1, \dots, d_n\}$ - კლასიფიცირებული ვებ-გვერდების კომპლექტი;

$A = D \times C$ - გადაწყვეტილების მატრიცა.

გადაწყვეტილების მატრიცის სტრუქტურას აქვს სახე:

| ვებ- გვერდები D_i | კატეგორიები - C_j | | | | |
|---------------------------|---------------------|-----|----------|-----|----------|
| | c_1 | ... | c_j | ... | c_k |
| d_1 | a_{11} | ... | a_{1j} | ... | a_{1k} |
| • | | | | | |
| • | | | | | |
| • | | | | | |
| d_i | a_{i1} | ... | a_{ij} | ... | a_{ik} |
| • | | | | | |
| • | | | | | |
| • | | | | | |
| d_n | a_{n1} | ... | a_{nj} | ... | a_{nk} |

სადაც, თითოეული ჩანაწერი: $a_{ij} \left(1 \leq i \leq n, 1 \leq j \leq k \right)$ აღნიშნავს - ვებ- გვერდი

d_i ეკუთვნის კატეგორია C_j - ს (ან არ ეკუთვნის). თითოეული $a_{ij} \in [0,1]$, სადაც 1 მიუთითებს ვებ-გვერდი d_i ეკუთვნის კატეგორია C_j - ს და 0 მანიშნებს, რომ არ ეკუთვნის.

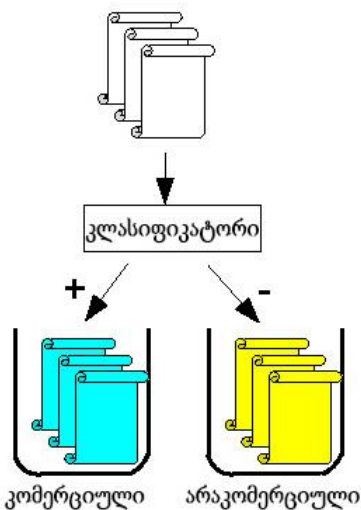
ვებ-გვერდი შეიძლება ერთ კატეგორიაზე მეტს მიეკუთვნებოდეს. ვებ-გვერდის კლასიფიცირების ამოცანა არის სავარაუდო უცნობი დავალების ფუნქცია $f: D \times C \rightarrow \{0,1\}$. სწავლის ფუნქციის საშუალებით $f: D \times C \rightarrow \{0,1\}$ ვუწოდებთ კლასიფიკატორს, მოდელს, სადაც f' ემთხვევა f -ს მაქსიმალურად [1].

ვებ-გვერდის კლასიფიკაციის პრობლემა შეიძლება დაიყოს რამდენიმე კატეგორიად, როგორცაა საგნობრივი კლასიფიკაცია, ფუნქციური კლასიფიკაცია, კლასიფიკაცია ობიექტზე ორიენტირებული მახასიათებლების მიხედვით და განწყობის კლასიფიკაცია [1].

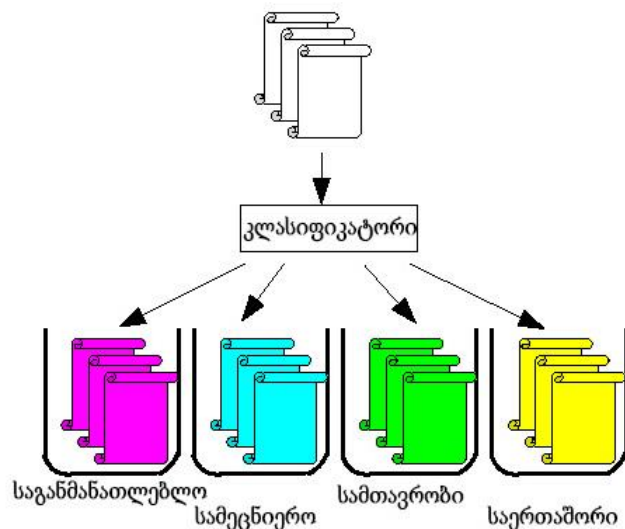
ჩვენ აქცენტს ვაკეთებთ ობიექტ-ორიენტირებული მახასიათებლების მიხედვით ვებ-გვერდის კლასიფიკაციაზე. სადაც განსაზღვრულია სხვადასხვა ობიექტის დეტალური მახასიათებლები და სწორედ ამ ნიშნის მიხედვით არის გამიჯნული ვებ-გვერდები ერთმანეთისგან.

3. ექსპერიმენტული ნაწილი

თემატიკაზე დაფუძნებული კლასიფიკაცია უკავშირდება ვებ-გვერდის საგანს. გვერდის სათაური შეიძლება იყოს „საგანმანათლებლო დაწესებულებები“, „სამეცნიერო“, „საერთაშორისო“, „ბიზნესი“, „ტექნოლოგიები“, „სპორტი“ და სხვ. მრავალი მახასიათებლის მიხედვით შეიძლება კლასიფიცირდეს ვებ-გვერდები, მაგრამ ამ ეტაპზე ვყოფთ ორ კატეგორიად: ორობითი და მრავალკლასობრივი კლასიფიკაცია (ნახ.1, 2).



ნახ.1. ორობითი კლასიფიკაცია



ნახ.2. მრავალკლასობრივი კლასიფიკაცია

ორობითი კლასიფიკაციის კატეგორია ზუსტად ქმნის ვებ-გვერდს პირველ ან მეორე კლასში. ხოლო მრავალკლასობრივი კლასიფიკაცია ქმნის ორზე მეტ კლასში. მაგალითად, ორობითი კლასიფიკაციის პრობლემის შემთხვევაში, ვებ-გვერდი შეიძლება მხოლოდ კლასიფიცირებული იყოს როგორც „კომერციული“ ან „არაკომერციული“, ხოლო მრავალკლასობრივი კლასიფიკაციის პრობლემის შემთხვევაში ვებ-გვერდი შეიძლება კლასიფიცირდეს ერთ კლასში შემდეგიდან: „საგანმანათლებლო“, „სამეცნიერო“, „სამთავრობო“, „საერთაშორისო“ და სხვ. (ნახ.3).



ნახ.3. კლასების დიაგრამა

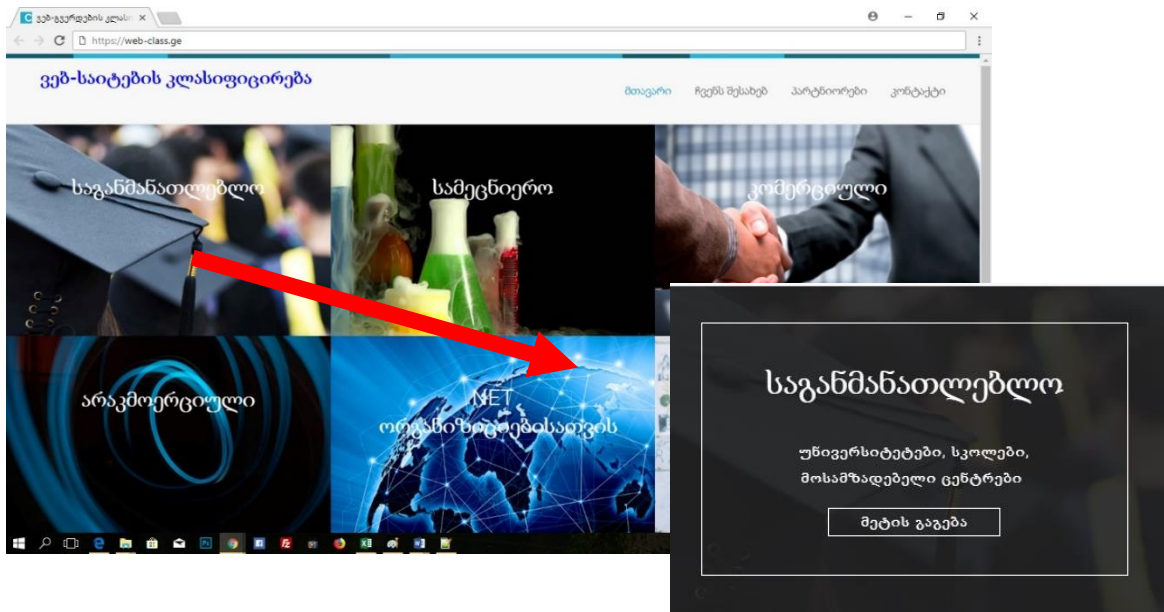
ინფორმაციის მოძიებისას მომხმარებლის ურთიერთობა ინტერნეტთან ხდება უფრო სრულყოფილი, როდესაც ერთი კლასის ქვეშ არსებული ვებ-გვერდები გაერთიანებულია ერთ სივრცეში. ამ დროს უმჯობესდება ძიების ხარისხი და გაცილებით მოკლე დროში ხდება შესაძლებელი იგივე შედეგის მიღება, რომელიც შესაძლოა დროში გაწელილიყო. ინტერნეტ სივრცეში არსებული ვებ-გვერდები დაყვავით ობიექტზე ორიენტირებული მახასიათებლების მიხედვით კლასებად.

სამუშაო პროცესმა გვაჩვენა, რომ მხოლოდ კლასებად დაყოფა არ არის საკმარისი და ამიტომ შემოვიღეთ ცნება ქვეკლასები. თითოეული კლასი ჩავშალეთ ქვეკლასებად, რის შედეგადაც მივიღეთ უფრო ზედმიწევნითი ერთიანი სისტემა. ამის საფუძველზე შევქმენით ერთიანი სივრცე (ვებ-გვერდი), სადაც ობიექტ-ორიენტირებული მახასიათებლების მიხედვით კლასიფიცირებულია ვებ-გვერდები. ვებ-გვერდის მთავარ გვერდზე გამოტანილია კლასები: „საგანმანათლებლო“, „სამეცნიერო“ და სხვ. (ნახ.4).

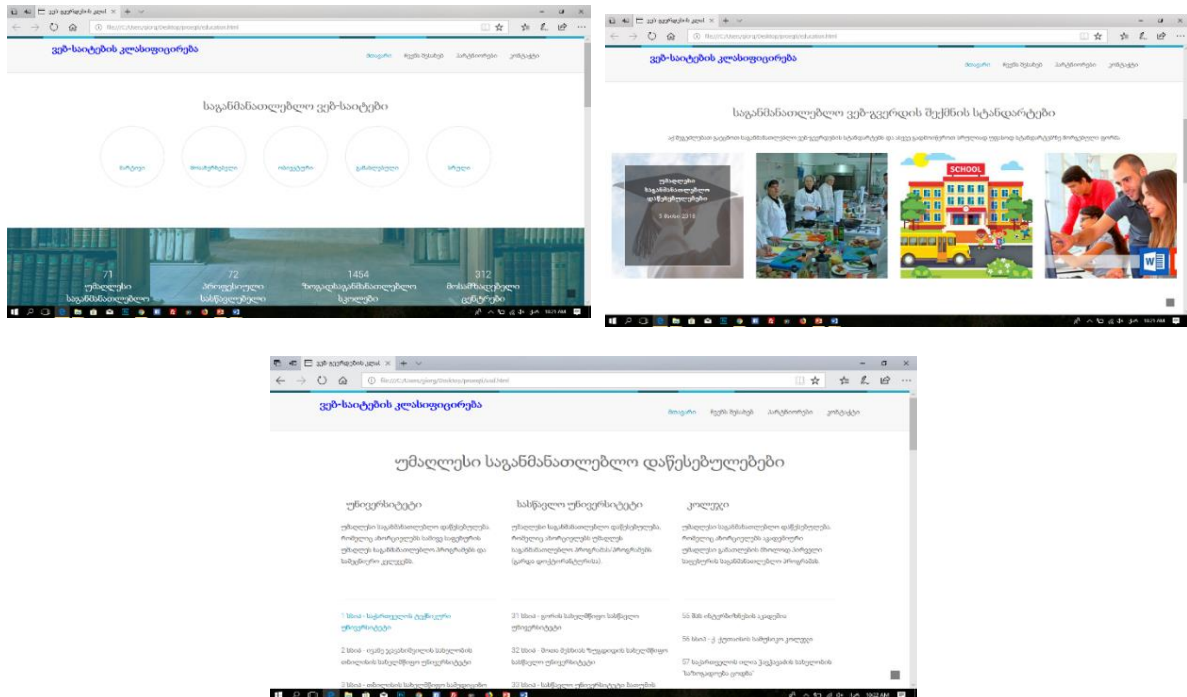


ნახ.4. ქვეკლასების დიაგრამა

აღნიშნული ვებ-გვერდი არის მრავალფუნქციური. მომხმარებელს შეუძლია მისი საშუალებით გაიგოს სხვადასხვა კლასის სტანდარტები, სურვილისამებრ გადმოიწეროს ამ კლასის შესაფერისი ვებ-შაბლონი და მისი საშუალებით ააწყოს საკუთარი ვებ-გვერდი. ასევე მისი ერთ-ერთი ღირებულებაა ინფორმაციულობა, ჩვენ შეგვიძლია ერთ სივრცეში მოვიძიოთ სხვადასხვა კლასის ქვეშ გაერთიანებული ვებ-გვერდები რომელიც უკვე ფუნქციონირებს ინტერნეტ სივრცეში (ნახ.5, 6).



ნახ.5. კლასიფიცირებული ვებ-გვერდის მთავარი გვერდი



ნახ.6. კლასიფიცირებული ვებ-გვერდის შიგა გვერდები

ლიტერატურა - References – Литература:

1. Qi X., Davison, B. (2009). Web page classification: Features and algorithms. ACM Comput. Surv., 41(2)
2. Choi B., Yao Z. (2008). Web Page Classification. Louisiana Tech. University, Ruston, LA 71272, USA, vol. 180, pp. 221-224
3. Zhengyang Lu. (2015). Web Page Classification Using Features from Titles and Snippets. Thesis of Master of Science in Electronic Business Technologies. University of Ottawa Ottawa, Ontario, Canada. https://ruor.uottawa.ca/bitstream/10393/33177/1/Lu_Zhengyang_2015_thesis.pdf

EXAMPLE OF CLASSIFICATION OF AN OBJECT-ORIENTED WEB SITE IN ACCORDANCE WITH IT'S MAIN CHARACTERISTICS

Kakashvili Giorgi
Georgian Technical University

Summary

Hence the fact that the web-pages of the world network have not been classified at the relevant level, which is confirmed with the conducted researches, we think that the information displayed on the web-pages should be organized with the strictly defined regulation. In the suggested model, which is reviewed in the article, the main accent is focused on the information function, value and sensitivity during the classification. There is worked out the classification of the web-page, as a problematic task, and there are defined the categories of web-page. The decision-making matrix is built, which defines the type of web-page.

ПРИМЕР КЛАССИФИКАЦИИ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ВЕБ-САЙТА В СООТВЕТСТВИИ С ЕЕ ОСНОВНЫМИ ХАРАКТЕРИСТИКАМИ

Какашвили Г.
Грузинский Технический Университет

Резюме

Исходя из того, что в мировой практике представленные сайты не классифицированы на начальном уровне, что подтверждают проведенные нами исследования. Мы считаем, что информационный поток, который расположен на веб-страницах должен приводится в порядок по строгой закономерности. По представленной нами модели на примере котором мы увидим, что когда происходит класификация веб-страниц, главный акцент направлен на значимость, ценность и сенситивность информации, которая расположена на веб-страницах. Также разработана класификация веб-страниц, как проблемная задача и определена категория веб-страниц. Построена матрица решений, с помощью которой определяются классы веб-страниц.