

Р. Ш. Гасанова

ОЦІНКА НАУКОВИХ СТАТЕЙ ЗА ДОПОМОГОЮ НОВОГО АЛГОРИТМУ ARTICLERANK

У статті розроблено нову модифікацію алгоритму PageRank для оцінки наукових статей і побудовано модель нового типу ArticleRank. У певних випадках для автоматичної оцінки наукових статей, розміщених в одній мережі, рекомендовано комп'ютерну програму мовою Delphi з широкими можливостями для проведення аналізу якісної оцінки наукових статей. Показано, що всередині однієї мережі статті, у яких однакова кількість посилань, мають різні значення ArticleRank.

Ключові слова: PageRank, ArticleRank, імпакт-чинник, індекс Хірша.

Вступ

Раніше основним показником наукової діяльності вченого вважалась кількість наукових статей. Останнім часом у зв'язку зі значним збільшенням кількості опублікованих наукових статей виникла необхідність їхньої якісної оцінки. Для цієї мети раніше існували різні системи оцінки [1]. За сучасного розвитку комп'ютерних технологій особливої актуальності набувають розробки передових наукоємних оцінок із практичним використанням комп'ютерної техніки. Загалом, останніми роками було розроблено різні механізми автоматичної оцінки наукової діяльності вчених і самих учених. Прикладом цього є імпакт-чинник – числовий показник наукового журналу, щороку обчислюваний Інститутом наукової інформації з 1960-го року. На сьогодні всередині корпорації Томсона є база даних, де зареєстровано більше, ніж 8400 наукових журналів різних країн світу [2]. Окрім цього, можна показати H-index для оцінки науковця, запропонований американським фізиком Джорджем Хіршем у 2005-ому році [3 – 6]. Цей індекс ґрунтується на кількості публікацій і цитат на кожну з цих публікацій. На сьогодні список показника H-index (індекс Хірша) для вчених, що працюють у різних галузях науки, представлений у реферативній базі даних Scopus of Web of Science [7]. Процес оцінки дисертаційних робіт із збереженням об'єктивності також вимагає від учених багато зусиль і часу. З метою автоматизації цього процесу деякі аспекти оцінки за якісними даними, такими як: визначення близькості між розділами в дисертації і визначенні актуальності роботи – описано в роботі [8]. Крім того, під час пошуку в Інтернеті поява на екрані як потрібної, так і непотрібної інформації засвідчує необхідність у фільтрації інформаційних сайтів. При цьому важливо, щоб найважливіші й найкорисніші інформаційні сайти розташовувалися в перших рядках списку, отриманого в результаті пошуку. Ця проблема вперше була розв'язана в 1998-ому році Лоуренсом Пейджем і Сергієм Бріном. Розроблений ними алгоритм, названий PageRank, успішно був випробуваний у системі Google і призвів до формування функціональної системи, у якій лише введенням url-адреси можна обчислити PageRank цього сайту [9]. Це програмне забезпечення, на якому ґрунтується пошукова система Google, збільшує ефективність процесу пошуку інформації. У математичному описі цього алгоритму використовують формулу, що показує рівень важливості – PageRank сторінки

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in B(p_i)} \frac{PR(p_j)}{O(p_j)},$$

де p_1, p_2, \dots, p_N – дані сторінки, $B(p_i)$ – число сторінок, які посилаються на сторінку p_i , $O(p_j)$ – число посилань, що виходять зі сторінки p_j , N – загальна кількість сторінок,

$d \in [0,8; 1]$ – коефіцієнт загасання.

Метою цієї статті є розробка нової модифікації алгоритму PageRank для оцінки наукових статей. Для досягнення цієї мети тут використано алгоритм PageRank із застосуванням розробок [10, 11], а також модель, що відображає важливість статті серед статей у конкретній галузі науки

$$AR(u) = \frac{1-d}{N} + d \sum_{v \in B(u)} \frac{AR(v)}{N_v}, \quad (1)$$

де $B(u)$ – множина статей, що посиляються на статтю u ; N_v – загальна кількість посилань, що виходять зі статті v ; d – коефіцієнт загасання моделі нового типу ArticleRank з додаванням у другий доданок зваженого коефіцієнта:

$$AR(u) = \frac{1-d}{N} + d \cdot w \cdot \sum_{v \in B(u)} \frac{AR(v)}{N_v}, \quad (2)$$

де $w = N_{vu} \cdot IF(v)$ – ваговий коефіцієнт; N_{vu} – кількість посилань від статті v до статті u ; $IF(v)$ – імпаکت-чинник журналу, у якому опубліковано статтю v .

Схематичний опис моделі ArticleRank. Розглянемо оцінювання статті, описаної в схемі (рис. 1):

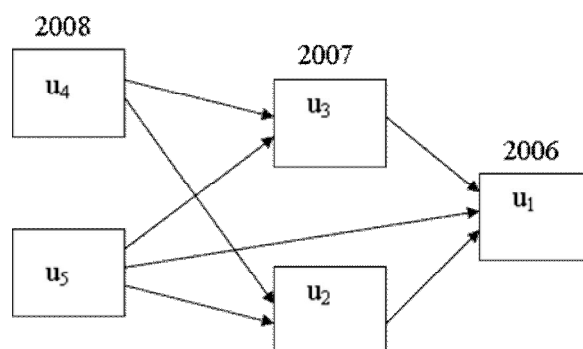


Рис. 1. Схематичний опис моделі ArticleRank

де u_2, u_3, u_4, u_5 – статті, що впливають на статтю (тобто множина статей, які посилалися на цю статтю безпосередньо або непрямо).

Як видно зі схеми, на першому етапі ArticleRank статті, що належить до останнього року, береться одиниця як в алгоритмі PageRank, оскільки на неї поки що немає посилань:

$$AR(u_4) = AR(u_5) = 1.$$

на другому етапі з формули (1) обчислюють ArticleRank статтей u_2 і u_3 відповідно

$$AR(u_2) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_4)}{2} + \frac{AR(u_5)}{3} \right) = 0,03 + 0,7055 = 0,7355;$$

$$AR(u_3) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_4)}{2} + \frac{AR(u_5)}{3} \right) = 0,03 + 0,7055 = 0,7355.$$

Отже, в цьому прикладі третій етап дає нам шукане значення оцінювання статті u_1 :

$$AR(u_1) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_5)}{3} + \frac{AR(u_2)}{1} + \frac{AR(u_3)}{1} \right) = 1,834.$$

Опишемо це оцінювання з додаванням пропонованих вагових коефіцієнтів. Для цього Наукові праці ВНТУ, 2012, № 1

введемо запропоновані вагові коефіцієнти у формулу (2) і схему.

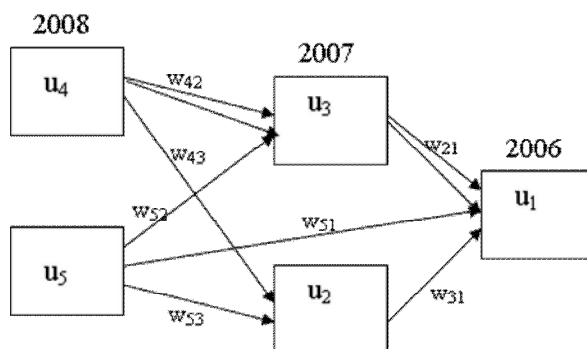


Рис. 2. Додавання в модель ArticleRank вагових коефіцієнтів

Використаємо значення імпаکت-чинника журналів, де опубліковано статті, описані у схемі як вагові функції. Припустимо, що журнали описані у схемі мають імпакт-чинники написані нижче:

$$IF(u_2) = 0,31, \quad IF(u_3) = 0,71, \quad IF(u_4) = 0,6, \quad IF(u_5) = 0,096$$

Імпакт-чинник статті u_1 поки що не потрібний, оскільки він не бере участь в оцінюванні самої статті. Присвоїмо ці значення статтям, описаним в останній схемі (рис. 2):

$$w_{42} = IF(u_4) \cdot N_{u_4} = 0,6 \cdot 2 = 1,2; \quad w_{43} = IF(u_4) \cdot N_{u_4} = 0,6 \cdot 1 = 0,6;$$

$$w_{51} = IF(u_5) \cdot N_{u_5} = 0,096 \cdot 1 = 0,096; \quad w_{52} = IF(u_5) \cdot N_{u_5} = 0,096 \cdot 1 = 0,096;$$

$$w_{53} = IF(u_5) \cdot N_{u_5} = 0,096 \cdot 1 = 0,096; \quad w_{21} = IF(u_2) \cdot N_{u_2} = 0,31 \cdot 2 = 0,62;$$

$$w_{31} = IF(u_3) \cdot N_{u_3} = 0,71 \cdot 1 = 0,71.$$

Після обчислення зважених функцій можна перейти до обчислення ArticleRank статті. Значення статті, опубліковані у 2008-ому році дорівнюють одиниці:

$$AR(u_4) = AR(u_5) = 1.$$

На другому етапі обчислюють:

$$AR(u_2) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_4)}{N_{u_4}} \cdot w_{42} + \frac{AR(u_5)}{N_{u_5}} \cdot w_{52} \right) = 0,03 + 0,85 \cdot \left(\frac{1}{3} \cdot 1,2 + \frac{1}{3} \cdot 0,096 \right) = 0,3972;$$

$$AR(u_3) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_4)}{N_{u_4}} \cdot w_{43} + \frac{AR(u_5)}{N_{u_5}} \cdot w_{53} \right) = 0,03 + 0,85 \cdot \left(\frac{1}{3} \cdot 0,6 + \frac{1}{3} \cdot 0,096 \right) = 0,2272;$$

$$AR(u_1) = \frac{1-0,85}{5} + 0,85 \cdot \left(\frac{AR(u_2)}{N_{u_2}} \cdot w_{21} + \frac{AR(u_5)}{N_{u_5}} \cdot w_{51} + \frac{AR(u_3)}{N_{u_3}} \cdot w_{31} \right) = 0,03 + 0,85 \times$$

$$\times \left(\frac{0,3972}{2} \cdot 0,62 + \frac{1}{3} \cdot 0,096 + \frac{0,2272}{1} \cdot 0,71 \right) = 0,299.$$

моделі на мові Delphi написана комп'ютерна програма, що дозволяє вибрати взаємозв'язані статті всередині бази й автоматичне обчислення ArticleRank. Основні дані програмних забезпечень задаються таким чином: у вікні "Adding" (рис. 3) потрібно ввести назви статей і їхніх посилань за зростанням років.

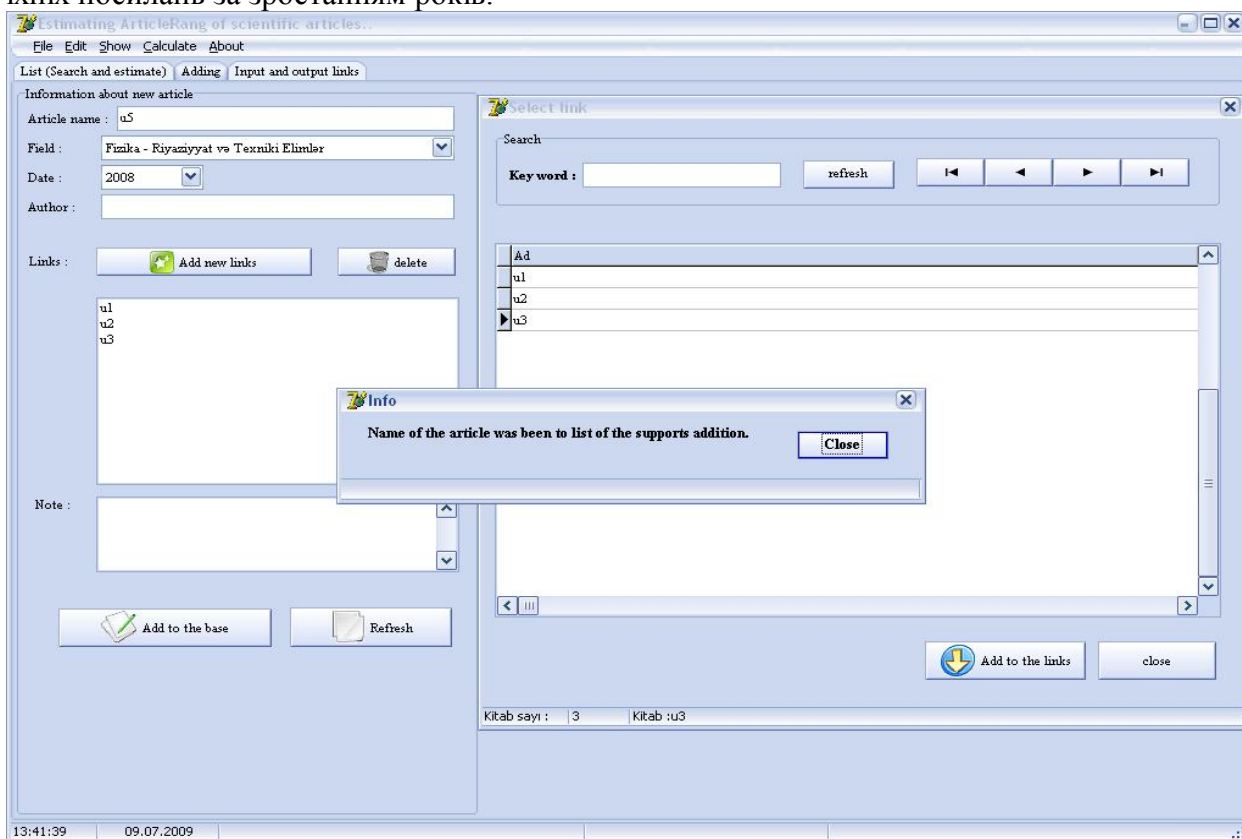


Рис. 3. Додавання статей і посилань у базу

Текст цього фрагмента на мові Delphi викладають таким чином:

```

procedure TEsas.TntBitBtn7Click(Sender: TObject);
Var
  i:integer;
begin
  If (TntEdit3.Text<>"")and(TntEdit4.Text<>"")and(TntComboBox3.Text<>"") then
  Begin
    TntBitBtn7.Enabled:=False;
    Esas.AdoQueryh.SQL.Text:='INSERT INTO siyahi(ad,tarix,deqiqlik,muellif'+
      ',elave) VALUES(:p0,:p1,:p2,:p3,:p4)';
    Esas.AdoQueryh.Parameters.ParamByName('p0').Value:=TntEdit3.Text;
    Esas.AdoQueryh.Parameters.ParamByName('p1').Value:=TntEdit4.Text;
    Esas.AdoQueryh.Parameters.ParamByName('p2').Value:=TntComboBox3.ItemIndex;
    Esas.AdoQueryh.Parameters.ParamByName('p3').Value:=TntEdit7.Text;
    Esas.AdoQueryh.Parameters.ParamByName('p4').Value:=TntMemo1.Text;
    Esas.AdoQueryh.ExecSQL;
    If TntListBox1.Items.Count>0 then
      For i:=0 to TntListBox1.Items.Count-1 do
      Begin
        AdoQueryh.SQL.Text:='INSERT INTO istinadlar(kid,iid) VALUES(:p0,:p1)';
        AdoQueryh.Parameters.ParamByName('p0').Value:=NewRecordNo;
        AdoQueryh.Parameters.ParamByName('p1').Value:=TntListBox2.Items.Strings[i];

```

```

AdoQueryh.ExecSQL;
End;
ElaveEdildi;
SiyahiniYaz;
TntBitBtn7.Enabled:=True;
ShowInfo('Info',Sozler.Siyahi.Lines.ValueFromIndex[0]);
End else
ShowInfo('Info',Sozler.Siyahi.Lines.ValueFromIndex[1]);
end;

```

У вікні “Input an output links” (рис. 4) відображують вхідні і вихідні посилання вибраної статті.

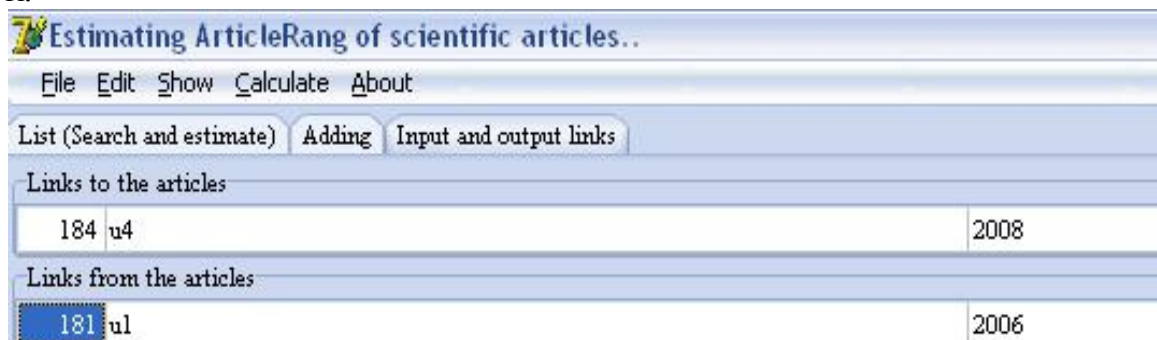


Рис. 4. Список вхідних і вихідних посилань із статті

Текст програми:

```

Procedure ButunCedvel(Id:integer);

```

```

Begin

```

```

Esas.ADOQuery4.SQL.Text:='SELECT Id,Ad,Tarix,Sbk,Elave FROM siyahi WHERE '+'
'(id in( SELECT iid FROM istinadlar WHERE '+'kid =:p0));'

```

```

Esas.ADOQuery4.Parameters.ParamByName('p0').Value:=id;

```

```

Esas.ADOQuery4.Open;

```

```

Esas.ADOQuery3.SQL.Text:='SELECT Id,Ad,Tarix,Sbk,Elave FROM siyahi WHERE '+'
'(id in( SELECT kid FROM istinadlar WHERE '+'
'iid =:p0));'

```

```

Esas.ADOQuery3.Parameters.ParamByName('p0').Value:=ID;

```

```

Esas.ADOQuery3.Open;

```

```

Esas.TntDBGrid3.Columns[0].Width:=40;

```

```

Esas.TntDBGrid3.Columns[1].Width:=400;

```

```

Esas.TntDBGrid3.Columns[2].Width:=90;

```

```

Esas.TntDBGrid3.Columns[3].Width:=50;

```

```

Esas.TntDBGrid3.Columns[4].Width:=100;

```

```

Esas.TntDBGrid4.Columns[0].Width:=40;

```

```

Esas.TntDBGrid4.Columns[1].Width:=400;

```

```

Esas.TntDBGrid4.Columns[2].Width:=90;

```

```

Esas.TntDBGrid4.Columns[3].Width:=50;

```

```

Esas.TntDBGrid4.Columns[4].Width:=100;

```

```

End;

```

У вікні List (Search and estimate) (рис. 5) можна обчислити ArticleRank статті й водночас відредагувати назву статті чи інші дані.

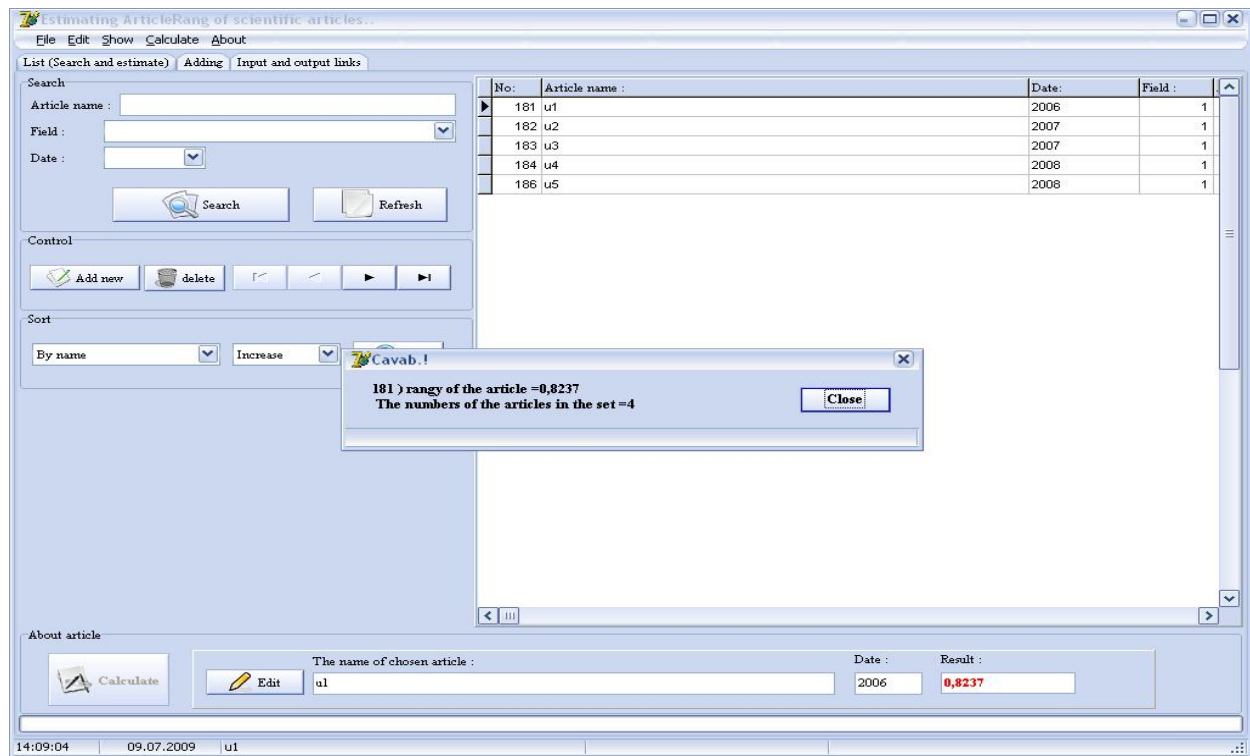


Рис. 5. Обчислення ArticleRank виділеної статті

Текст програми цього вікна буде таким:

```
Function RandomName:String;
Var
  s:string[11];
Begin
  Randomize;
  s:=chr(random(26)+65)+chr(random(26)+65)+chr(random(26)+65)+chr(random(26)+65)+chr(random(26)+65)+c
hr(random(26)+65);
  RandomName:=s;
End;
Procedure SebekeleriYenile;
Begin
  Esas.TntPageControl1.Pages[0].Show;
  Esas.ADOQueryH.SQL.Text:='Select id,tarix from siyahi Order by tarix asc';
  Esas.ADOQueryH.Open;
  Esas.ADOQueryH.First;
  SebekeniYenileN:=Esas.ADOQueryH.RecordCount;
  SebekeniYenileI:=0;
  Esas.TntProgressBar1.Max:=SebekeniYenileN ;
  Esas.TntProgressBar1.Min:=SebekeniYenileI;
  Esas.SebekeniYenileT.Enabled:=True;
End;
//*****//
Function SebekdekiJurnallar(id:integer):integer;
Var sbk:string[10];
Begin
  Esas.ADOQueryH.SQL.Text:='SELECT sbk FROM siyahi WHERE id=:p0';
  Esas.ADOQueryH.Parameters.ParamByName('p0').Value:=id;
  Esas.ADOQueryH.Open;
```

```

sbk:=Esas.ADOQueryH.Fields.Fields[0].Value;
Esas.ADOQueryH.SQL.Text:='SELECT id FROM siyahi WHERE sbk=:p0';
Esas.ADOQueryH.Parameters.ParamByName('p0').Value:=Sbk;
Esas.ADOQueryH.Open;
SebekdekiJurnallar:=Esas.ADOQueryH.RecordCount;
End;
// *****//
//           Estimating           //
// *****//
Procedure MassiviYenile;
Var
i:integer;
Begin
For i:=0 to MaxCount do
count[i]:=0;
End;
Function AR(n,k,id:integer):real;
Var
c,rar:real;
j:byte;
Begin
Esas.ADOQueryH.SQL.Clear;
Esas.ADOQueryH.SQL.Text:='SELECT id,Ad FROM siyahi WHERE '+'
'id in( SELECT kid FROM istinadlar WHERE '+'
'id =:p0)';
Esas.ADOQueryH.Parameters.ParamByName('p0').Value:=id;
Esas.ADOQueryH.Open;
Esas.ADOQueryH.First;
count[k]:=Esas.ADOQueryH.RecordCount;
If count[k]>0 then
Begin
For j:=0 to count[k]-1 do
Begin
data[k,j].id:=Esas.ADOQueryH.Fields.Fields[0].AsInteger;
data[k,j].icount:=IstinadSayi(Esas.ADOQueryH.Fields.Fields[0].AsInteger);
Esas.AdoQueryH.Next;
End;
c:=0;
For j:=0 to count[k]-1 do
c:=c+ar(n,k+1,data[k,j].id)/data[k,j].icount;
rar:=(1-d)/n+d*c;
End else
Begin
rar:=1;
End;
Ar:=rar
End;
Procedure Hesabla;
Var
rAR:real;
sAR:string;

```

```

n:integer;
begin
If Esas.ADOQuery1.Fields.Fields[0].AsString<>" then
Begin
Esas.HesablaB.Enabled:=False;
n:=SebekdekiJurnallar(Esas.AdoQuery1.Fields.Fields[0].Value);
Esas.TntPageControl1.Pages[0].Show;
MassiviYenile;
rAR:=AR(n,0,Esas.AdoQuery1.Fields.Fields[0].AsInteger);
sAR:=Copy(FloatToStr(rAR),1,6);
Esas.ARText.Text:=sAR;
ShowInfo('Cavab.!',Esas.AdoQuery1.Fields.Fields[0].AsString+Sozler.Siyahi.Lines.Strings[5]+sAR+#13#10+Sozler.Siyahi.Lines.Strings[8]+IntToStr(n));
Esas.HesablaB.Enabled:=True;
End;
End;

```

Висновки

На основі представлених моделей та їхньої програмної реалізації отримані такі результати:

1. Аналогічно алгоритму PageRank для оцінки web-сторінок розроблено алгоритм ArticleRank для оцінки наукових статей.

2. Введено в модель і ефективно використано вагові функції, що визначають посилання з інших статей на аналізовану статтю й розкрито їхній вплив на обчислення ArticleRank.

3. У певних випадках для автоматичної оцінки наукових статей, які розташовані в одній мережі, рекомендована комп'ютерна програма мовою Delphi з широкими можливостями для проведення аналізу якісної оцінки наукових статей.

4. Усередині однієї мережі статті, що мають однакові кількості посилань мають різні значення ArticleRank, оскільки кожне посилання має власний ваговий коефіцієнт.

СПИСОК ЛІТЕРАТУРИ

1. Бернал Дж. Наука в истории общества. / Дж. Бернал. – М., 1956. – 311 с.
2. Thomson Reuters [Електронний ресурс] // Режим доступу: <http://www.thomsonreuters.com>.
3. Hirsch J. E. An index to quantify an individual's scientific research output / J. E. Hirsch // Proceedings of the National Academies of Science. – 2005. – Vol. 102. – No. 46. – P. 16569 – 16572.
4. Hirsch, J. E. (2007). Does the h-index have predictive power? [Електронний ресурс] // Режим доступу: http://arxiv.org/PS_cache/arxiv/pdf/0708/0708.0646v2.pdf.
5. Bar-Ilan J. Which h-index? / J. Bar-Ilan // A comparison of WoS, Scopus and Google Scholar Scientometrics. – 2008. – Vol. 74. – No. 2. – P. 257 – 271.
6. Cronin B. Using the h-index to rank influential information scientists / B. Cronin, L. I. Meho // Journal of the American Society for Information Science and Technology. – 2006. – Vol. 57. – No. 9. – P. 1275 – 1278.
7. Jacso Peter. The pros and cons of computing the h-index using Google Scholar? / Peter Jacso // Online. Information Review. – 2008. – Vol. 32. – No. 3. – P. 437 – 452.
8. Гасанова Р. Ш. Автоматизація оцінювання дисертаційних робіт / Р. Ш. Гасанова // Вісник ВПІ. – 2009. – № 1. – С. 89 – 91.
9. Этичное продвижение сайтов. [Електронний ресурс] // Режим доступу: <http://linkclub.ru>.
10. Brin S. The anatomy of a large-scale hypertextual web search engine / S. Brin, L. Page // Computer Networks ISDN Systems. – 1998. – vol. №№ 1 – 7. – P. 107 – 117.
11. Ma N. Bringing PageRank to the citation analysis / N. Ma, J. Guan, Y. Zhao // Information Processing and Management. – 2008. – vol. 44. – № 2 – P. 800 – 810.

Гасанова Рахіля Шабан кизи – науковий співробітник інституту інформаційних технологій Національної академії наук Азербайджана (ІТ НАНА), тел.: (994 12) 4397213, e-mail: rahasanova@gmail.com.
Національна академія наук Азербайджану.