

Р. Ш. Гасанова

ОЦЕНКА НАУЧНЫХ СТАТЕЙ С ПОМОЩЬЮ НОВОГО АЛГОРИТМА ARTICLERANK

В статье разработана новая модификация алгоритма PageRank для оценки научных статей и построена модель нового типа ArticleRank. В определенных случаях для автоматической оценки научных статей, размещенных в одной сети, рекомендована компьютерная программа на языке Delphi с широкими возможностями для проведения анализа качественной оценки научных статей. Показано, что внутри одной сети статьи, у которых одинаковое количество ссылок, имеют различные значения ArticleRank.

Ключевые слова: PageRank, ArticleRank, импакт-фактор, индекс Хирша.

Введение

Раньше основным показателем научной деятельности учёного было принято считать количество научных статей. Последнее время в связи со значительным увеличением количеств опубликованных научных статей возникла необходимость их качественной оценки. Ранее для этой цели существовали различные системы оценки [1]. При современном развитии компьютерных технологий особую актуальность приобретают разработки передовых наукоёмких оценок с практическим использованием компьютерной техники. В последние годы были разработаны различные механизмы автоматической оценки научной деятельности ученых и самих ученых. Примером этого является импакт-фактор – числовой показатель научного журнала, ежегодно вычисляемый Институтом научной информации с 1960-го года. На сегодняшний день внутри корпорации Томсона имеется база данных, в которой зарегистрированы более 8400 научных журналов различных стран [2]. Кроме этого, можно привести H-index для оценки научного работника, предложенный американским физиком Джорджем Хиршем в 2005-ом году [3 – 6]. Этот индекс основан на количестве публикаций и ссылок на каждую из этих публикаций. В данный момент список показателя H-index (индекс Хирша) для ученых, работающих в различных областях науки, представлен в реферативной базе данных Scopus of Web of Science [7]. Процесс оценки диссертационных работ с сохранением объективности также требует от ученых много усилий и времени. С целью автоматизации этого процесса некоторые аспекты оценки по таким качественным данным, как: определение близости между разделами в диссертации и определении актуальности работы – описаны в работе [8]. Кроме того, во время поиска в Интернете появление на экране как нужной, так и ненужной информации свидетельствует о необходимости фильтрации информационных сайтов. При этом важно, чтобы самые важные и полезные информационные сайты располагались в первых строках списка, полученного по результатам поиска. Эта проблема была впервые разрешена в 1998-ом году Лоуренсом Пейджем и Сергеем Брином. Разработанный ими алгоритм, названный PageRank, был успешно испытан в системе Google и привел к формированию функциональной системы, в которой только вводом URL-адреса можно вычислить PageRank этого сайта [9]. Это программное обеспечение, на котором основана поисковая система Google, увеличивает эффективность процесса поиска информации. В математическом описании этого алгоритма используют формулу, показывающую уровень важности – PageRank страницы p_i :

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

где p_1, p_2, \dots, p_N – рассматриваемые страницы, $B(p_i)$ – число страниц ссылающихся на страницу p_i , $O(p_j)$ – число ссылок, выходящих со страницы p_j , N – общее количество страниц, $d \in [0,8; 1]$ – коэффициент затухания.

Целью данной статьи является разработка новой модификации алгоритма PageRank для оценки научных статей. Для достижения этой цели здесь использован алгоритм PageRank с применением разработок [10, 11], а также модель, отражающая важность статьи u среди статей в конкретной области науки

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

где $B(u)$ – множество статей, ссылающихся на статью u ; N_v – общее количество ссылок, выходящих из статьи v ; d – коэффициент затухания модели нового типа ArticleRank с добавлением во второе слагаемое взвешенного коэффициента:

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

где $w = N_{vu} \cdot IF(v)$ – весовой коэффициент; N_{vu} – количество ссылок от статьи v к статье u ; $IF(v)$ – импакт-фактор журнала, в котором опубликована статья v .

Схематическое описание модели ArticleRank. Рассмотрим оценивание статьи u_1 , описанной в схеме (рис. 1):

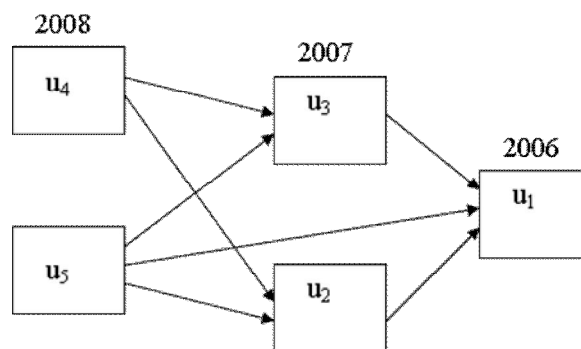


Рис. 1. Схематическое описание модели ArticleRank

где u_2, u_3, u_4, u_5 – статьи, влияющие на статью (т. е. множество статей, которые ссылались на эту статью непосредственно или косвенно).

Как видно из схемы, на первом этапе ArticleRank статьи, относящейся к последнему году, принимается за единицу, как и в алгоритме PageRank, поскольку на неё пока ещё нет ссылок:

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

На втором этапе по формуле (1) вычисляют ArticleRank статей u_2 и u_3 соответственно:

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

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

Следовательно, в этом примере третий этап дает нам искомое значение оценивания статьи u_1 :

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

Опишем это оценивание с добавлением предлагаемых весовых коэффициентов. Для этого включим предложенные весовые коэффициенты в формулу (2) и схему (рис. 2).

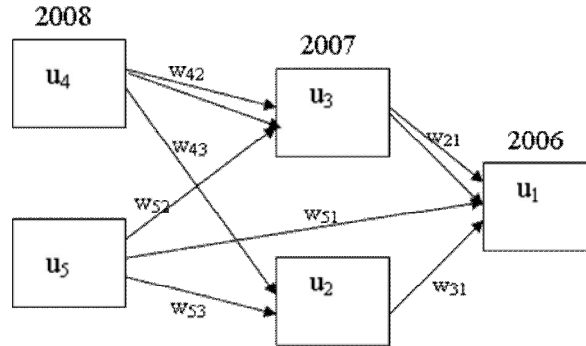


Рис. 2. Добавление в модель ArticleRank весовых коэффициентов

Используем значение импакт-фактора журналов, где опубликованы статьи, описанные в схеме как весовые функции. Допустим, что журналы, описанные в схеме, имеют написанные ниже импакт-факторы:

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

Импакт-фактор статьи u_1 пока что не нужен, поскольку он не участвует в оценивании самой статьи. Присвоим эти значения статьям, описанным в последней схеме:

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

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

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

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

$$w_{53} = IF(u_5) \cdot N_{u_5} = 0,096 \cdot 1 = 0,096;$$

$$w_{21} = IF(u_2) \cdot N_{u_2} = 0,31 \cdot 2 = 0,62;$$

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

После вычисления взвешенных функций можно перейти к вычислению ArticleRank статьи. Значения статьи, опубликованные в 2008-ом году равны единице:

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

На втором этапе вычисляют:

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

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

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

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

Программное обеспечение модели ArticleRank на языке Delphi. На основе построенной модели на языке Delphi написана компьютерная программа, обеспечивающая выбор взаимосвязанных статей в базе и автоматическое вычисление ArticleRank. Основные данные для программного обеспечения задаются следующим образом: в окне “Adding” (рис. 3) нужно ввести названия статей и их ссылок по возрастанию годов.

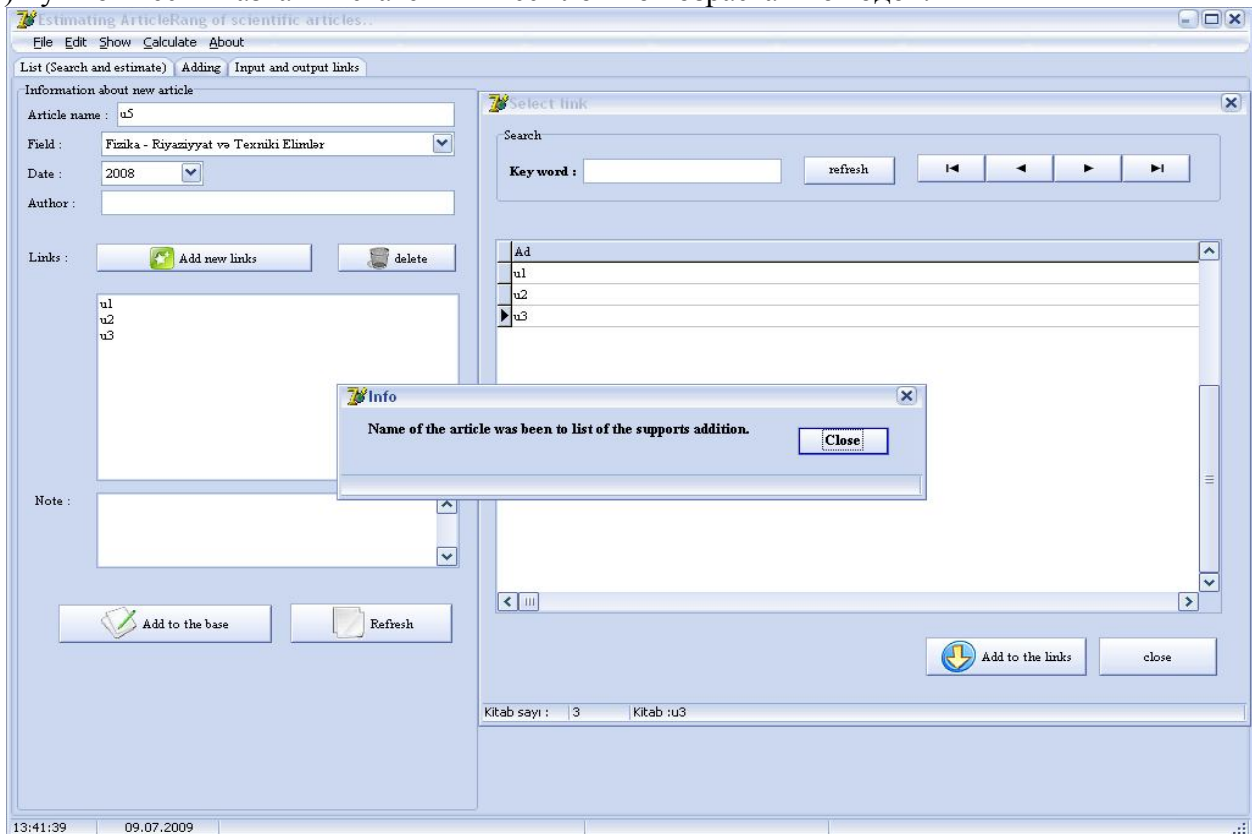


Рис. 3. Добавление статей и ссылок в базу

Текст этого фрагмента на языке Delphi излагают таким образом:

```
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;

```

В окне “Input an output links” (рис. 4) отображают входящие и исходящие ссылки выбранной статьи.

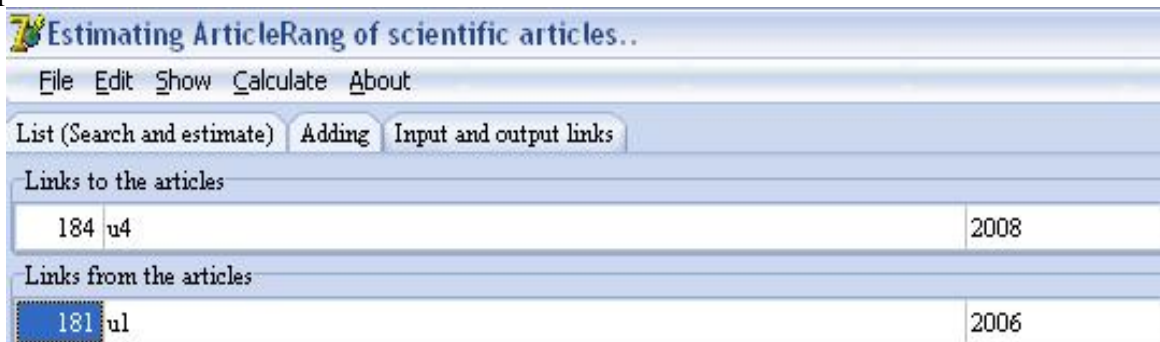


Рис. 4. Список ссылок, входящих и исходящих из статьи

Текст программы:

```

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;
```

В окне List (Search and estimate) (рис. 5) можно вычислить ArticleRank статьи и одновременно отредактировать название статьи или другие данные.

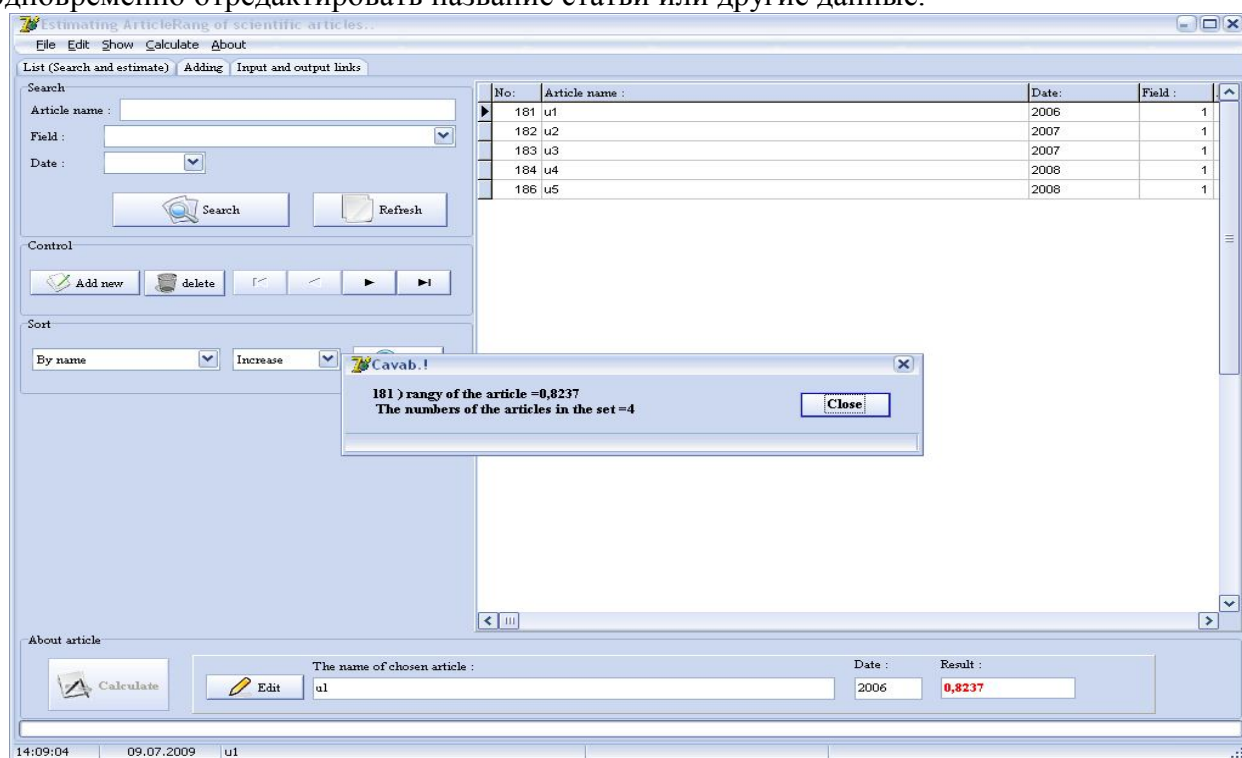


Рис. 5. Вычисление ArticleRank выделенной статьи

Текст программы этого окна будет такой:

```
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 '+'
    'iid =: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;

```

Выводы

На основе представленных моделей и их программной реализации получены следующие результаты:

1. Разработан алгоритм ArticleRank для оценки научных статей аналогичный алгоритму PageRank для оценки web-страниц.
2. Введены в модель и эффективно использованы весовые функции, определяющие ссылки из других статей на анализируемую статью, раскрыто их влияние на вычисление ArticleRank.
3. В определенных случаях для автоматической оценки научных статей, размещенных в одной сети, рекомендована компьютерная программа на языке Delphi с широкими возможностями для проведения анализа качественной оценки научных статей.
4. Внутри одной сети статьи, имеющие одинаковое количество ссылок, имеют различные значения ArticleRank, поскольку каждая ссылка имеет собственный весовой коэффициент.

СПИСОК ЛИТЕРАТУРЫ

1. Бернал Дж. Наука в истории общества. / Дж. Бернал. – М., 1956. – 311 с.
2. Thomson Reuters [Электронный ресурс] // Режим доступа: <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? [Электронный ресурс] // Режим доступа: 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. [Электронный ресурс] // Режим доступа: <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 Рахия Шабан кызы – научный сотрудник института информационных технологий Национальной академии наук Азербайджана (ИИТ НАНА), тел.: (994 12) 4397213, e-mail: rahasanova@gmail.com.

Национальная академия наук Азербайджана.