https://ravshanfikr.tj/nav-alamon/muajyankunii-muallifi-asar-oi-ta-rikhiyu-sijos-bo-vositai-trigramma-oi-ramz.html

УДК 81’322::811.222.8::519.25
ББК 81.1::Ш152.131.2-923::В172

МУАЙЯНКУНИИ МУАЛЛИФИ АСАРҲОИ ТАЪРИХИЮ-СИЁСӢ БО ВОСИТАИ ТРИГРАММАҲОИ РАМЗӢ

Қосимов А.А. – *номзади илмҳои техникӣ, омӯзгори калон, кафедраи барномарезӣ ва низомҳои иттилоотӣ, Донишкадаи политехникии Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ, abdunabi_kbtut@mail.ru*
Қаюмов А.А. – *магистрант, Академияи миллии илмҳои Тоҷикистон Маркази илмии Хуҷанд, zohirjon96@gmail.com*

Чакида. *Муқаррар карда шуд, ки басомади вохӯрии триграммаҳои забони тоҷикӣ дар асарҳои таърихӣ - сиёсӣ ин муайянкунандаи муаллиф аст. Барои муайянкунии муаллифи матн бо басомади вохӯрии ҳарфҳои триграмма ва имконияти истифодабарии таснифгари Усмонов З.Ҷ. таҳқиқот гузаронида шуд. Роҳҳо, мақсад ва имкониятҳои таснифгари Усмонов З.Ҷ. оварда мешаванд. Симои рақамӣ ва фазои ченаки асарҳои таърихӣ-сиёсӣ сохта шудаанд. Бо мақсади ягонагии эҷодиёти муаллиф, қимати ченаки муқоисакунанда муқаррар карда мешаванд, ки дар асоси он ягонагии синфҳои асарҳо муайян мешаванд. γ-таснифгари бузургии тасодуфии фосиладор, ки самаранокии баландро дар муайян кардани муаллифи асарҳои шоирону нависандагони форсу тоҷик нишон дода буданд, барои муайян кардани муаллифи асарҳои таърихӣ-сиёсӣ тафтиш карда мешаванд.*

Калидвожаҳо: *забони тоҷикӣ, асарҳои таърихӣ-сиёсӣ, триграмма, таснифгар, басомади вохӯрӣ.*

Натиҷаҳое, ки олимон дар ин ҷо ба даст овარданд, фақат барои асарҳои бадеӣ ва иқтисодӣ буданд. Аз ҳамин нуқтаи назар дар ин мақола маълумот оид ба таҳқиқоти муайян кардани муаллифи асарҳои таърихӣ-сиёсӣ бо воситаи басомади вохӯрии триграммаҳои рамзӣ оварда мешавад. Масъалаи шинохти матн, дар асл, аз он рӯзе, ки хат пайдо шуд, ба миён омадааст. Дар муддати тӯлонӣ, он танҳо бо яке аз “рӯя”-ҳои худ, ки дар он зарурати муайян кардани иҷрокунандаи асари навишташуда ифода меёфт, мавриди таҳлил ва баррасӣ қарор додбиграма мешуд. Минбаъд, пас аз лаҳзаи ихтироъ шудани “китобчопкунӣ”, дар ин масъала, рӯяи нави муҳими дигари он –

зарурати муайян кардани муаллифи маводи чопӣ ба миён омад. Дар марҳилаи кунунӣ маҳз ҳамин "рӯя"-и дуюм, мазмуни мундариҷаи асосии ин масъаларо ташкил медиҳад.

Дар кори зерин ба сифати инструменти таҳқиқотшаванда таснифгари Усмонов З.Ҷ. санҷида мешавад, [3-5].

Маълумот оид ба коллексияи матнҳо. Асарҳои муаллифҳои зерин гирифта шуданд ва дар дохили қафс шакли кӯтоҳ карда шудаи онҳо, ки барои ҷойгиркунии ҷадвалҳои 1 ва 2-и поёни лозим буд, оварда шудаанд: А. Дюма (АД) "Граф монте кристо 1" (ГМК1) ва " Граф монте кристо 2" (ГМК2) [6, 7], А. Хуш (АХ) "Шӯриши Усмон 1" (ШУ1) ва " Шӯриши Усмон 2" (ШУ2) [8-10], М. Шакурӣ (МШ) "Садри Бухоро" (СБ) ва "Хуросон аст инҷо" (Х) [11, 12], С. Айни (СА) "Ёддоштҳо 1" (Ё1) ва " Ёддоштҳо 2" (Ё2) [13, 14], Ҷ. Икромӣ (ҶИ) "Дувоздаҳ дарвозаи Бухоро" (ДБ) ва "Он чи аз сар гузашт" (О) [15, 16]. Ба омӯзиши ин масъала ҳамагӣ 5-муаллиф ва 10-асар гирифта шуд.

Таснифгари матнҳо. Ба сифати тавсири рақамии асарҳои таърихӣ-сиёсӣ басомади вохӯрии триграммаҳои рамзӣ дида баромада мешавад. Барои муайянкунии муаллиф як метод – таснифгари матнии Усмонов З.Ҷ. истифода бурда шуд. Моҳияти тавсифи ин метод дар татбиқ ба масъалаҳои илми забоншиносӣ дода мешавад, [3].

Бигзор T_1 ва T_2 – ду матне бошанд, ки қонуни тақсимоти триграммаҳои рамзии онҳо ба намуди ҷадвал дода шуда бошад

$$\begin{array}{ll} T_i: & 1 \;\cdots\; k \;\cdots\; m \\ P^{(i)}: & p_1^{(i)} \cdots p_k^{(i)} \cdots p_m^{(i)}, \end{array} \tag{1}$$

ки дар ин ҷо

$$\sum_{k=1}^{m} p_k^{(i)} = 1 \qquad \text{аст.}$$

Дар ин ифодаҳо k ($k = \overline{1,m}$) - рақами тартибии триграммаи k-юм дар алифбои триграмма, $p_k^{(i)}$- басомади нисбии вохӯрии триграммаи k-юм дар матни T_i, $i = 1, 2$ мебошад. Он гоҳ масофаи байни T_1 ва T_2 бо формулаи зерин муайян карда мешавад

$$\rho(T_1, T_2) = \sqrt{\frac{m}{2}} \max_{s} \left| \sum_{k=1}^{s} (p_k^{(1)} - p_k^{(2)}) \right|, \tag{2}$$

дар ин ҷо $s = \overline{1,m}$.

Бигзор γ - ягон адади мусбат бошад, матнҳои T_1 ва T_2 *γ-якҷинса* номида мешаванд, агар

$$\rho(T_1,\, T_2) \le \gamma. \tag{3}$$

ва *γ-ғайриякҷинса* номида мешаванд, агар

$$\rho(T_1,\, T_2) > \gamma \text{ бошад.} \tag{4}$$

Фарз мекунем, ки коллексияи матнҳо T ба зермаҷмӯъҳои $T^{(j)}$, $j = \overline{1, n}$ тақсим шудааст. Барои қиммати қайдшудаи γ адади $\aleph^0$- суммаи ҷуфтҳои якҷинсаи матн, ки ба зермаҷмӯъҳои $T^{(j)}, j = \overline{1, n}$, тааллуқ доранд ва адади $\aleph^н$-суммаи γ- ҷуфтҳои ғайриякҷинса, ки ба зермаҷмӯъҳои гуногун тааллуқ дорад, ҳисоб карда мешавад. Нисбати

$$\eta = \frac{\aleph^0 + \aleph^н}{\mathbb{N}}, \tag{5}$$

ки дар ин ҷо $\mathbb{N}$-шумораи умумии ҷуфти матнҳо дар коллексияи T аст, барои қиммати дода шудаи γ самаранокии татбиқи модели математикии (1) – (4) ба таври автоматӣ тақсимкунии коллексияи T ба зерқисми $T^{(j)}$-ро тавсир мекунад. Дар мақолаҳои [4, 5], барои ҳисоб кардани қимати оптималии $\gamma^{опт}$, ки барои он самаранокии максималии η барои коллексияи $T = \{T^{(j)}\}$ дастрас мегардад, пешниҳод гардид.

Натиҷаҳо. Алгоритми дар боло зикр шударо истифода бурда, комплекси барномаҳо тартиб дода шуданд ва дар аввал басомади вохӯрии триграммаҳои рамзӣ бе ва бо ҳисобгирии фосила дар алоҳидагӣ ҳисоб карда, баъдан масофаи байни асарҳо бо формулаи (2) ҳисоб карда шуданд, натиҷаҳо дар ҷадвалҳои 1 ва 2 оварда шудаанд. Аз натиҷаҳои ба даст омада чунин қонуниятро бояд ҷудо кард, ки ду асари як муаллиф якҷинсаанд ва ду асари муаллифашон гуногун ғайриякҷинсаанд.

https://ravshanfikr.tj/nav-alamon/muajyankunii-muallifi-asar-oi-ta-rikhiyu-sijos-bo-vositai-trigramma-oi-ramz.html

Ҷадвали 1

Қиммати $\gamma^{опт}$ барои триграмма бе ҳисобгирии фосила

АД		АХ		МШ		СА		ЧИ	
ГМК1	**ГМК2**	**ШУ1**	**ШУ2**	**СБ**	**Х**	**Ё1**	**Ё2**	**ДБ**	**О**
0,1861									
0,6390	**0,5833**								
0,4607	**0,4050**	**0,4365**							
0,6781	**0,5994**	**0,4274**	**0,3602**						
1,0023	**0,9127**	**0,4249**	**0,6016**	**0,4307**					
0,3541	**0,2840**	**0,7365**	**0,6183**	**0,7991**	**1,1212**				
0,2818	**0,2134**	**0,6563**	**0,5422**	**0,7059**	**0,9656**	**0,1896**			
0,3944	**0,3419**	**0,3959**	**0,3208**	**0,4310**	**0,7058**	**0,4742**	**0,3687**		
0,3726	**0,3186**	**0,3747**	**0,3092**	**0,3819**	**0,6916**	**0,4732**	**0,3713**	**0,2413**	

Бояд қайд кард, ки дар ҳар ду ин ҷадвал дар диагонали асосӣ маълумот оид ба муносибати байни асарҳои як муаллиф, аммо дар дигар ячейкаҳо маълумот оид ба муносибати байни асарҳои муаллифашон гуногун оварда шудаанд.

Барои муайянкунии муаллиф диапазони қиммати мувофиқи γ бо воситаи триграмма бе ҳисобгирии фосила баробари γ = [0,2134; 0,2413) шуд. Дар ин ҳолат бо нобаробарии зерин

$$\rho(T_1, T_2) \leq [0{,}2134;\ 0{,}2413) \qquad (6)$$

- якҷинсагии ҷуфти асарҳо, аммо бо муқобили нобаробари,

$$\rho(T_1, T_2) > [0{,}2134;\ 0{,}2413), \qquad (7)$$

- бо ғайриякҷинсагии асарҳо мувофиқат мекунад. Ин қоидаро ба қатори ададҳои ҷадвали 1 татбиқ намоем, нобаробарии (6) дар 16 ячейка риоя намешавад, аммо нобаробарии (7) фақат дар 2 ячейка рия намешавад. Ба ҳолати зерин, таъсирнокии метод бо формулаи (5) ҳисоб карда шуд, ки баробари η = 91% аст.

https://ravshanfikr.tj/nav-alamon/muajyankunii-muallifi-asar-oi-ta-rikhiyu-sijos-bo-vositai-trigramma-oi-ramz.html

Ҷадвали 2

Қиммати γ^{onm} барои триграмма бо ҳисобгирии фосила

АД		АХ		МШ		СА		ЧИ	
ГМК1	**ГМК2**	**ШУ1**	**ШУ2**	**СБ**	**Х**	**Ё1**	**Ё2**	**ДБ**	**О**
0,1317									
0,4466	**0,3240**								
0,5435	**0,4218**	**0,1575**							
0,4233	**0,4947**	**0,7271**	**0,7727**						
0,5595	**0,6315**	**0,7968**	**0,9006**	**0,3379**					
0,2150	**0,3066**	**0,5550**	**0,6157**	**0,3945**	**0,4952**				
0,1497	**0,2211**	**0,4301**	**0,5394**	**0,3355**	**0,4336**	**0,1445**			
0,1257	**0,2041**	**0,4919**	**0,5453**	**0,3995**	**0,5039**	**0,1781**	**0,1900**		
0,2636	**0,3350**	**0,5240**	**0,5714**	**0,3314**	**0,3919**	**0,1496**	**0,1261**	**0,1837**	

Таҳлили ҷадвали 2 нишон медиҳад, ки бо воситаи триграмма бо ҳисобгирии фосила, таснифгар ҳиссиёти баландтарро доро аст. Барои муайянкунии муаллиф диапазони қиммати мувофиқи γ бо воситаи триграмма бо ҳисобгирии фосила баробари γ = [0,1257; 0,3379) шуд. Дар ин ҳолат бо нобаробарии зерин

$$\rho(T_1,\ T_2) \leq [0{,}1257;\ 0{,}3379) \qquad (8)$$

- якҷинсагии ҷуфти асарҳо, аммо бо муқобили нобаробари,

$$\rho(T_1,\ T_2) > [0{,}1257;\ 0{,}3379), \qquad (9)$$

- бо ғайриякҷинсагии асарҳо мувофиқат мекунад. Ин қоидаро ба қатори ададҳои ҷадвали 2 татбиқ намоем, нобаробарии (8) дар 10 ячейка риоя намешавад, аммо нобаробарии (9) бошад, фақат дар 2 ячейка рияо намешавад. Ба ҳолати зерин, таъсирнокии метод бо формулаи (5) ҳисоб

карда шуд, ки баробари η = 89% аст.

Хулоса. Аз маълумотҳои ҳангоми таҳқиқот ба даст омада, ба чунин хулосаҳо омадан мумкин аст, ки

- триграммаи рамзӣ дар масъалаи муайянкунии муаллифи матни таърихӣ-сиёсӣ ба сифати тавсифҳои миқдорӣ комилан қобили қабул мебошанд;

- ба ҳисобгирии фосила дар триграммаҳо саҳеҳии таснифотро баланд мебардорад;

- таснифгари Усмонов З.Ҷ. (1) – (5) дараҷаи кифоя калони муайянкунии муаллифи матнҳои таърихӣ-сиёсиро нишон медиҳад.

ЛИТЕРАТУРА

1. Усманов З.Д., Косимов А.А. Частотность триграмм в таджикской литературе.– Доклады Академии наук Республики Таджикистан, 2016, т.59, № 1-2, с. 28-32.

2. Романов А.С., Шелупанов А.А., Мещеряков Р.В. Разработка и исследование математических моделей, методик и программных средств информационных процессов при идентификации автора текста.– Томск: -В-Спектр, 2011, 188 с.

3. Усманов З.Д. N-граммы в распознавании однородных текстов.– Материалы 20 научно-практического семинара "Новые информационные технологии в автоматизированных системах".– Москва, 2017, с. 52-54.

4. Усманов З.Д. Классификатор дискретных случайных величин.– Доклады Академии наук Республики Таджикистан. 2017, т.60, № 7-8, с. 291-300.

5. Усманов З.Д. Алгоритм настройки кластеризатора дискретных случайных величин.– Доклады Академии наук Республики Таджикистан, 2017, т.60, № 9, с. 392-397.

6. Александр Дюма Граф монте кристо. 1844, 520 с.

7. Алӣ Хуш Шӯриши Усмон . Душанбе, 2008, 128 с.

8. Муҳаммадҷон Шакурӣ Садри Бухоро. Душанбе, 2005, 330 с.

9. Муҳаммадҷон Шакурӣ Хуросон аст инҷо.

10. Садриддин. Айнӣ Ёддоштҳо. 1990, 352 с.

11. Ҷалол. Икромӣ Дувоздаҳ дарвозаи Бухоро. 1969, 474 с.

12. Ҷалол. Икромӣ Он чи аз сар гузашт. Душанбе, 2009, 360 с.

https://ravshanfikr.tj/nav-alamon/muajyankunii-muallifi-asar-oi-ta-rikhiyu-sijos-bo-vositai-trigramma-oi-ramz.html

ИДЕНТИФИКАЦИЯ АВТОРОВ ИСТОРИЧЕСКОЙ ПОЛИТИЧЕСКОЙ ПРОИЗВЕДЕНИЙ С ПОМОЩЬЮ СИМВОЛНЫХ ТРИГРАММ

Косимов А.А. – *кандидат технический наук, старший преподаватель, кафедра программирования и информационных технологий, Политехнический институт Таджикского технического университета имени академика М.С. Осими, abdunabi_kbtut@mail.ru*

Каюмов А.А. – *магистрант, Академия наук Республики Таджикистан, Худжандский научный центр, zohirjon96@gmail.com*

Аннотация. *Устанавливается, что распределение частотности триграмм в историческых-политическых произведениях таджикского языка является идентификатором авторства. Исследованы возможности классификатора З.Д.Усманова распознавать автора текста по частотности буквенных триграмм. Сконструированы цифровой портрет и метрическое пространство произведений. В предположении уникальности авторского творчества устанавливаются пороговые значения метрики, на основе которых определяются классы "однородных" произведений. -классификатор дискретных случайных величин, подтвердивший высокую эффективность при идентификации авторства текстовых фрагментов в произведениях классической и современной поэзиях, а также в современной прозе таджикского языка, тестируется на предмет приспособляемости к распознаванию авторства в исторической -политической произведений.*

Ключевые слова: *таджикский язык, исторический-политический произведения, триграмма, классификатор, частотность.*

IDENTIFICATION OF AUTHORS OF A ECONOMICS-POLITICAL TEXT BY MEANS OF A SYMBOL TRIGRAMS

Kosimov A.A. – *Candidate of technical Sciences, Department of Programming and Information Technologies, Polytechnic Institute of Tajik Technical University, abdunabi_kbtut@mail.ru*

Kayumov A.A. – *master student, National Academy of sciences of Tajikistan, Khujand scientific cente, zohirjon96@gmail.com*

https://ravshanfikr.tj/nav-alamon/muajyankunii-muallifi-asar-oi-ta-rikhiyu-sijos-bo-vositai-trigramma-oi-ramz.html

Annotation. *It is established that the distribution of the frequency of trigrams in the historical and political works of the tajik language is the identifier of authorship. The possibilities of the classifier Z.D.Usmanov to recognize the author of the text by the frequency of letter trigrams are investigated. A digital portrait and a metric space of works are constructed. Assuming the uniqueness of the author's creativity, the threshold values of the metrics are established, on the basis of which the classes of "homogeneous" works are determined. The γ-classifier of discrete random variables, which confirmed high efficiency in identifying the authorship of text fragments in works of classical and modern poetry, as well as in modern prose of the Tajik language, is tested for adaptability to the recognition of authorship in historical-political text.*

Key words: tajik language, *historical-political text, trigram, classifier, frequenc.*