

E-dövlət Sisteminin Analizində Data Mining Texnologiyalarının Tətbiq İmkanları

Ramiz Aliquliyev¹, Günay Niftəliyeva²

^{1,2}AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹r.aliguliyev@gmail.com; ²gunayniftali@gmail.com

Xülasə — Big data dövlətin qarşısında duran yeni və mühüm məsələlərdən biridir. Big data analitikləri e-dövlətin effektiv və səmərəli idarə olunmasında text və data mining metodlarının mühüm rol oynadığını qeyd edirlər. Bunu nəzərə alaraq, məqalədə bu sahədə müxtəlif tədqiqatçılar tərəfindən görülmüş işlər nəzərdən keçirilmiş, text və data mining texnologiyalarının rolu müəyyən olunmuşdur.

Açar sözlər — e-dövlət; big data; data mining; text mining.

I. GİRİŞ

Elektron dövlət dörd dəyişənin funksiyası kimi müəyyən edilə bilər: idarəetmə, informasiya və kommunikasiya texnologiyaları (İKT), biznes proseslərinin yenidən təşkili və e-vətəndaş. Eləcə də buraya, rəqəmsal media vasitəsilə vətəndaşlara, biznes və ya digər dövlət orqanlarına milli və dövlət informasiya və xidmətlərinin göstərilməsi daxildir [1]. E-dövlət quruculuğu dövlətlə vətəndaş arasında informasiya mübadiləsi və öhdəliklərin qarşılıqlı olaraq "rəqəmsal mühitdə" davamlı, etibarlı, şəffaf bir şəkildə reallaşdırılmasını ifadə edir. İKT-nin sürətli inkişafı e-dövlətin inkişafına kömək edir. Vətəndaşların gündəlik həyatında İKT-dən istifadənin artması bu sahədə e-dövlət xidmətlərinin genişlənməsinə səbəb olmuşdur. Ədəbiyyatlarda "E-dövlət sistemləri dövlətə inam, sosial inteqrasiya, ictimai rifah kimi siyasi və sosial vəzifələri yerinə yetirmək üçün effektivlik, səmərəlilik və iqtisadiyyat həddini aşan strateji məqsədləri əhatə edir" kimi diqqətə çatdırılır [2].

Big data, bulud hesablamalar e-dövlətin son inkişaf meyilləri kimi göstərilir. Big data müəyyən zaman aralığında müxtəlif mənbələrdən toplanan faydalı və istifadə edilə bilən elektron məlumatların böyük həcmidir. Mənbələrdə qeyd olunur ki, "dünya məlumatlarının 90%-i son iki il ərzində toplanmışdır və hər gün 2.5 kvintilyon bayt məlumat əlavə edilir" [2]. Bu məlumatların böyük həcmi öz strukturuna görə relyasion verilənlər bazasında istifadə edilə bilməz. Verilənlər yeni dəyər kəsb edən qiymətli mənbələrdir. Dövlətin vətəndaşlara xidmət göstərməsi, səhiyyə xərcləri, işsizlik, təbii fəlakətlər və terrorizmin artması kimi milli problemlərin aradan qaldırılmasında big data-dan istifadə oluna bilər. Bu yaxınlarda, B. Obama "rəqəmsal məlumatların böyük, mürəkkəb kolleksiyalarından bilik, məlumat əldə etmək imkanlarının yaxşılaşdırılması; elm və texnikanın kəşf tempini sürətləndirmək üçün bu texnologiyalardan istifadə edilməsi; milli təhlükəsizliyin möhkəmləndirilməsi; tədris və təlimi transformasiya etmək üçün" Big Data Tədqiqat və İnkişaf Təşəbbüsünə (Big Data Research and Development Initiative) başlamışdır [2]. Big data bir sıra digər texnologiyaların tətbiqi

ilə yanaşı, e-dövlətdən istifadəni genişləndirməyə kömək edə bilər. Verilənlər eksponensial sürətlə artdığından ənənəvi məlumat (data) emalı vasitələri və texnologiyaları bu məlumatların analizində faydalı ola bilmir. Big data analitikləri bunun text və data mining texnologiyaları vasitəsilə mümkün olduğunu qeyd edirlər [1, 3].

II. ƏLAQƏLİ İŞLƏRİN İCMALI

E-dövlətdə data mining - dövlət veb saytlarında verilənlərin (məlumatların) qərar qəbul etmədə istifadə oluna biləcək faydalı biliyə çevrilməsi prosesidir. E-dövlətdə text və data mining texnologiyalarının tətbiqi ilə bağlı bir sıra tədqiqatlar aparılmışdır.

[4]-də mətn məlumat mənbələrindən bilik və əlaqələrin tapılmasında text mining texnologiyasından istifadə olunmasının faydaları müzakirə olunmuşdur. Bu məqalədə dövlət siyasəti və vətəndaş rəyləri arasında əlaqəni aşkar etmək üçün veb forum və bloqlarda text mining metodlarından istifadə olunması və vətəndaş xidmətlərinin bununla təkmilləşdirilə biləcəyi təklif olunur. Bundan başqa, çoxdillilik text mining, maşın tərcümə (machine translation), ikidillilik text mining-in texniki arxitekturası və s. kimi mühüm məqamlar müzakirə olunur.

[5]-də dövlət qərarlarının qəbul olunmasında daha yaxşı nəticələrin əldə olunması üçün data mining texnologiyasının rolu müzakirə olunur. Bu məqalədə e-dövlətin qurulmasında data mining metodu tətbiq olunur. İlk növbədə ictimai rəy araşdırılır, daha sonra SPSS və korrelyasiya analizi vasitəsilə toplanan məlumatlar emal olunur, insanların real ehtiyaclarına xas qayda tapılır və beləliklə dövlət qərarı üçün daha yaxşı dəstək təmin olunur.

[6]-da e-dövlətdə açıq məlumatların (verilənlərin) analizi ilə bağlı yanaşma təklif olunub. Bu yanaşmaya 4 məsələ daxildir: məlumatların toplanması və emalı, modelin aşkar olunması və modelin təhlili.

[7]-də elektron səsvermə üçün konseptual model və e-səsvermə vasitəsilə toplanmış məlumatlar üzərində data mining texnologiyasının müxtəlif üstünlükləri müzakirə olunmuşdur. Bu məqalədə data mining texnologiyasının e-səsvermə verilənlər bazasından biliyin aşkarlanmasında çox səmərəli olduğu göstərilmişdir.

E-dövlət üçün qeyri-səlis data mining alqoritmi [8]-də təklif olunur. Burada e-dövlət məsələləri sistemində faydalı məlumatı aşkarlamaq üçün qeyri-səlis data mining-dən istifadənin faydaları müzakirə olunur. Sonda bu qənaətə gəliblər ki, e-

dövlətdə qeyri-səlis data mining-dən istifadə siyasətçilərə böyük kömək ola bilər.

[9]-da e-dövlətdə data mining texnologiyasının tətbiqi üçün yanaşma təklif olunur. Bu məqalədə e-dövlətdə data mining texnologiyasının tətbiqi 2 kateqoriyaya bölünür. Birinci kateqoriya ümumi tətbiqdır və bura vətəndaşların identifikasiyası, insan resurslarının və layihələrin idarə olunması daxildir. İkinci kateqoriyaya kənd təsərrüfatı, səhiyyə, təhsil və s. daxildir. Bu tədqiqatlarda həmçinin dövlət təşkilatları daxilində qərarların qəbulu üçün data mining-dən istifadə olunaraq müxtəlif məsələlər, problemlər araşdırılmışdır. Sonda bu qənaətə gəlməmişlər ki, data mining texnologiyası təkcə saxtakarlıq (dələduzluq) və təhlükəsizlik sahəsində təhdidləri aşkar etməyə kömək etmir, həmçinin dövlət-dövlət, dövlət-vətəndaş, dövlət-biznes kimi e-dövlət xidmətlərinin təkmilləşməsinə təsir edərək vətəndaşların davranış və istəkləri kimi təsiredici faktların ölçülməsində istifadə oluna bilər.

[10]-da vətəndaşların rifahı üçün e-dövlətin tətbiq və xidmət imkanlarının genişləndirilməsi üçün text və data mining texnologiyalarının tətbiqi təklif olunmuşdur. Bu məqalədə text və data mining texnologiyalarından ictimai rəyin modelləşdirilməsi və hökumət qərarlarının qiymətləndirilməsində istifadə olunur. Daha sonra ictimai rəyin istiqamətini əldə etmək və digər tərəfdən əlaqədar məsələlər üzrə yeni hökumət qərarlarının hazırlanması və çıxarılan rəylər arasında əsas korrelyasiyanı müəyyən etmək üçün vətəndaş rəyləri analiz olunur.

E-dövlət xidmətlərinin təkmilləşdirilməsi üçün [11]-də hibrid yanaşma təklif olunur. Burada, e-dövlət sənədlərinin çoxsəviyyəli klasterləşməsi (multilayered clustering) məqsədilə müxtəlif mətn yaxınlıq ölçüləri və qeyri-səlis konsepsiyanın tətbiqi təklif olunur.

[12]-də veb forumlarda vətəndaşların mesaj və rəylərinin analizi vasitəsilə səsvermə tendensiyalarını müəyyən etmək üçün text mining texnologiyasına əsaslanan yanaşma təklif olunmuşdur.

Hökumət qərarlarına ictimaiyyətin mövqeyini müəyyən etmək üçün [13]-də yanaşma irəli sürülmüşdür. Bu yanaşmada text mining-dən istifadə edərək əvvəlcə dövlət qərarları haqqında vətəndaş rəyləri toplanır və daha sonra analiz olunur. Text mining metodları mətn verilənlərindən xüsusiyyətlərin çıxarılması nəticəsində yaradılan strukturlaşdırılmış verilənlər bazasında istifadə olunur. Burada mətn verilənləri K-Means alqoritmi vasitəsilə biliyin aşkarlanması üçün klasterləşdirilir. Strukturlaşdırılmış mətn verilənlərindən əldə olunan bilik e-dövlət üzrə strateji qərarların dəstəklənməsində istifadə oluna bilər.

III. E-DÖVLƏT SİSTEMİNDƏ TEXT VƏ DATA MİNING TEXNOLOGİYALARININ İSTİFADƏ İMKANLARI

E-dövlətdə İnformasiya Texnologiyalarının (İT) rolu elmə yaxşı məlumdur. Verilənlər bazasından biliyin praqmatik istifadəsi, biliyin idarə olunması texnologiyaları e-dövlət mühitində qərar qəbulətmə sistemlərinə böyük töhfə verə bilər. E-dövlətdə müxtəlif təbiətli (mətn, video, audio) məlumatlar

toplanır. Aydın ki, e-dövlətin effektiv idarə olunmasında bu məlumatların analizi mühüm əhəmiyyət kəsb edir. Bu məlumatların həcmi çox böyük sürətlə artır və bu e-dövlət mühitində böyük həcmli verilənlərin analizi probleminə gətirib çıxarır. Hal-hazırda verilənlərin effektiv analizində data mining texnologiyaları çox geniş tətbiq olunur. Data mining verilənlər bazasında biliyin aşkar olunmasında ən mühüm texnologiyalardan biridir və biliklərin idarə edilməsində əhəmiyyətli sahə hesab olunur. Data mining verilənlər bazasından gizli, amma faydalı biliyin aşkarlanmasını həyata keçirir. Verilənlər bazasından biliyin aşkarlanmasının formal tərifini aşağıdakı kimi verilir: Data mining verilənlər haqqında dolayısı ilə əvvəllər məlum olmayan və potensial faydalı informasiyanın çıxarışıdır. Aşkar edilmiş bilik e-dövlət sistemində xidmətin keyfiyyətinin yaxşılaşdırılmasında istifadə oluna bilər. Data mining metodları və bilik idarəetmə texnologiyaları birlikdə biliklə zəngin mühit yaratmaq üçün tətbiq edilə bilər. Verilənlər bazasında biliyin aşkarlanması böyük həcmli verilənlərlə işlədikdə strateji qərarları inkişaf etdirmək və modellərin müəyyən olunmasında effektiv ola bilər. [14, 15].

Qeyd edək ki, data mining metodları marketing, maliyyə və səhiyyə sahəsində də uğurla tətbiq edilir. Dövlət cinayət və terror fəaliyyətinin, dələduzluq hallarının aşkar olunması, göstərilən xidmətlərin yaxşılaşdırılması, elmi məlumatların təhlili, insan resurslarının idarə olunmasında data mining-dən istifadə edə bilər. Təşkilatlar əhalinin ehtiyaclarını ödəmək üçün nəzərdə tutulmuş siyasətin planlaşdırılması, həyata keçirilməsi və qiymətləndirilməsi üçün data mining texnologiyasını tətbiq edir. Bütün bu sahələrdə planlaşdırma və qərar qəbulətmənin keyfiyyəti dəqiq obyektiv məlumatların toplanması, təhlili və interpretasiyasından asılıdır [16, 17].

E-dövlət mühitində məlumatların çox böyük sürətlə artdığını və bu məlumatların əksəriyyətinin mətn şəklində saxlanıldığını nəzərə alsaq, text mining metodlarının da data mining-lə yanaşı yüksək potensiala malik olduğunu qeyd edə bilərik. Text mining statistika, maşın təlim, məlumat-axtarış, data mining, linqvistik və təbii dil emalı anlayışlarını bir araya gətirən nisbətən yeni fənlərə sahədir. Bu avtomatik olaraq müxtəlif yazılı mənbələrdən informasiyanın çıxarılması ilə yeni, əvvəllər məlum olmayan informasiyanın kompüter tərəfindən aşkar olunması prosesidir. Text mining bir sıra xüsusiyyətlərinə görə data mining-ə oxşardır. Data mining əsasən rəqəmli strukturlaşdırılmış məlumatlarla, text mining isə struktursuz məlumatlarla işlədikdə faydalıdır [4].

E-dövlətdə istifadə olunan text mining metodları bunlardır: informasiyanın axtarışı, informasiyanın çıxarılması, mövzunun izlənilməsi, referatlaşdırma, təsnifat, konsepsiya (anlayış) əlaqələri, informasiyanın vizuallaşdırılması və sualın cavablandırılması.

İnformasiyanın çıxarılması: İnformasiyanın çıxarılması alqoritmləri mətn daxilində əsas ifadələri və əlaqələri müəyyən edir. Bu "model uyğunluğu (şablon)" adlandırılan prosesdən istifadə edərək mətdə əvvəlcədən müəyyən edilmiş ardıcılıqların axtarılması ilə həyata keçirilir. Alqoritmlər istifadəçini mənalı fikir ilə təmin etməkdən ötrü bütün müəyyən edilmiş ardıcılıqlar arasında əlaqələri aşkarlayır. Bu metod böyük həcmli mətn ilə işlədikdə çox faydalıdır.

Təsnifat: Təsnifat sənədin verilmiş mövzular çoxluğuna yerləşdirilməsi ilə onun əsas mövzusunun müəyyən olunmasını nəzərdə tutur. Burada informasiyanın çıxarılması metodunda olduğu kimi aktual informasiyanın emal olunmasına cəhd göstərilir. Təsnifat yalnız mətdə olan sözləri sayır və bu say sənədin əhatə olunduğu əsas mövzuları müəyyən edir. Təsnifat tez-tez mövzuların əvvəlcədən müəyyən olduğu lüğətə (thesaurus) əsaslanır və əlaqələr geniş terminlər, dar terminlər, sinonimlər və müvafiq terminlərə görə müəyyən olunur.

Klasterləşmə: Klasterləşmə oxşar sənədləri qruplaşdırmaq üçün istifadə olunan metoddur. Burada təsnifatdan fərqli olaraq əvvəlcədən təyin olunmuş mövzulardan istifadə etmək əvəzinə sənədlərin bir-birinə oxşarlığı əsasında qruplaşdırma aparılır. Əsas klasterləşmə alqoritmi hər bir sənəd üçün mövzuların vektorunu yaradır və sənədin hər bir klasterə nə dərəcədə uyğun gəldiyini ölçür.

Mövzunun izlənməsi: Mövzunun izlənmə sistemi istifadəçi profillərini istifadə etməklə işləyir, istifadəçi tərəfindən baxılan sənədlərə əsaslanır və istifadəçidə maraq doğuran digər sənədləri proqnozlaşdırır. Text mining vasitələrindən bəziləri istifadəçilərə maraq doğuran xüsusi kateqoriyaları seçməyə imkan verir və istifadəçinin marağını onun oxuma tarixi və klik məlumatları əsasında avtomatik olaraq müəyyən edir.

Referatlaşdırma: Mətnin əsas müddəalarını və ümumi mənasını saxlamaqla həcmnin azaldılmasından ibarətdir.

Sualın cavablandırılması: Text mining-in digər sahələrindən biri sualın cavablandırılmasıdır. Burada o verilmiş suala ən yaxşı cavabın necə tapılması ilə məşğul olur. Sualın cavablandırılmasında bir çox text mining metodlarından istifadə oluna bilər.

Assosiasiyaların aşkarlanması: Assosiasiya qaydaları mövzular arasında və ya əlaqəli mətn çoxluğunu xarakterizə etməkdə istifadə olunan təsviri anlayışlar arasında olan əlaqələrin və təsirlərin öyrənilməsində istifadə olunur.

Qeyd etdiyimiz bu metodlar text miningin həll etdiyi əsas məsələlərdir. Big data analitikləri bu metodların e-dövlətdə böyük həcmli məlumatların analizində mühüm rol oynayacağını qeyd edirlər.

NƏTİCƏ

E-dövlət İKT-dən istifadə etməklə dövlət və vətəndaşlar arasında münasibətlərə təsir edərək dövlət sektorunda xidmətlərin yaxşılaşdırılması üçün dövlət idarəçiliyində yeni mərhələ hesab olunur. E-dövlətdə toplanan məlumatların çox böyük sürətlə artması bu sahədə bir sıra problemlər yaradır. Bu problemlərə elektron mühitdə eksponensial sürətdə yaranan informasiya axınına analiz etmək üçün səmərəli üsulların inkişaf etdirilməsi tələbatı daxildir. Bu məlumatlar yalnız ölçülərinə görə böyük deyil, həm də dinamik və heterogen xarakterlidir. Yuxarıda göstərilən problemləri nəzərə alaraq tədqiqatçılar tərəfindən bu verilənlərdən biliyin aşkarlanması üçün text və data mining texnologiyalarının istifadə olunmasına baxmayaraq e-dövlətin daha yaxşı həyata keçirilməsi üçün daha səmərəli və etibarlı üsulların inkişaf etdirilməsinə ehtiyac var. Text və data mining texnologiyalarının e-dövlət mühitində həyata keçirilməsində əsas problemlər aşağıdakılardır:

- ✓ Verilənlər bazasının dinamik təbiəti
- ✓ Eksponensial sürətdə yaranan verilənlər bazasının böyük həcmi
- ✓ Verilənlər bazasının heterogen təbiəti
- ✓ Gizli verilənlər bazaları
- ✓ Atributların seçilməsi

Elektron məlumatların saxlanması problemləri böyük çətinliklər yaradır. Big data analitikanın inkişafı bu problemin öhdəsindən gələ bilər və e-dövlət layihələrinin səmərəli həyata keçirilməsinə kömək edə bilər.

ƏDƏBİYYAT

- [1] K.K.Verma, N. Shrivastava, “Exploring Role and Associated Challenges of Data Mining Technique and its Implementation in E-Governance”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 4, No. 8, pp. 3473-3478, 2015.
- [2] Joseph M. Woodside, Sh. Amiri, and B. Boldrin, “The Impact of ICT and Big Data on e-Government”, Proceedings of the Int'l Conf. on Advances in Big Data Analytics, pp. 117-118, 2015.
- [3] Y.-Ch. Chen, T.-Ch. Hsieh, “Big Data for Digital Government: Opportunities, Challenges, and Strategies”, International Journal of Public Administration in the Digital Age, Vol. 1, No.1, pp.1-14, 2014.
- [4] G. Koteswara Rao, Shubhamoy Dey, “Decision Support for E-Governance: A Text Mining Approach”, International Journal of Managing Information Technology (IJMIT) Vol.3, No.3, pp.73-91, 2011.
- [5] G. Guan, L. Zhou, & P. Tang, “Research of Data Mining in Government Transparent Decision-making”, The 2009 International Symposium on Web Information Systems and Applications (WISA 2009), pp. 335-338, 2009.
- [6] P. Milić, N.Veljković & L. Stoimenov, “Framework for open data mining in e-government”, Proceedings of the Fifth Balkan Conference in Informatics ACM, pp. 255-258, September 2012.
- [7] S. Kumar, H. Aggarwal, U. Kaur & Sumati, “E-Governance and Data Mining: a methodology for e-voting”, International Journal of Information Technology and Knowledge Management, Vol 2, No. 1, pp. 139-144, 2009.
- [8] A. Mujawar and V. Patil , “Fuzzy based data mining system for E-government”, Elixir Comp. Sci. & Engg. 51A, pp. 11066-11068, 2012.
- [9] V. R. Rao, “ A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services - An Indian Perspective”, International Journal of Computer Science and Information Technology Research, Vol. 2, No 4, pp. 209-225, 2014.
- [10] G. Stylios, D. Christodoulakis, J. Besharat, M.-A. Vonitsanou, I. Kotrotsos, A. Koumpouri, S. Stamou, “Data Mining Applications in the Service of E-Government”, Digital Democracy, Vol. 1, pp. 544-558, 2012.
- [11] G. Šimić, Z. Jeremić, E. Kajan, D. Randjelović, A. Presnall, “A Framework for Delivering e-Government Support”, Acta Polytechnica Hungarica, Vol. 11, No. 1, pp. 79-96, 2014.
- [12] J. Cardeñosa, C. Gallardo , J.M. Moreno-Jiménez, “Text mining techniques to support e-democracy systems”, Proceedings of the International Conference on E-Learning, E-Business, Enterprise Information System and E-Government, pp. 401-405, 2009.
- [13] A. B. Adeyemo and A. K. Ojo, “Classification of Social Blogs Comments Using Text Mining”, International Journal of Computer Science, Vol. 11, No 1, pp. 54-58, 2014.
- [14] G. Sangeetha, Dr. L.M. Rao, “A Review on Contribution of Data mining in e-Governance Framework”, International Journal of Engineering Research and General Science, Vol. 3, No. 2, pp. 68-75, 2015.
- [15] G. Sangeetha , R. L. Manjunatha, “Efficient Framework of e-Government for Mining Knowledge from Massive Grievance Redressal

- Data”, International Journal of Advanced Research in Computer and Communication Engineering ,Vol. 4, No. 5, pp. 671-677, 2015.
- [16] P. Desai, “A Survey of Data Mining Research and Implementation for E-Governance Systems”, International Journal of Computer Science, Vol. 4, No.7, pp. 103-105, 2014.
- [17] Dr. R. R. Velamala, “A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services” -An Indian Perspective”, International Journal of Computer Science and Information Technology Research, Vol. 2, No. 4, pp. 209-225, 2014.