



TANQIDIY NAZAR, TAHLILIIY TAFAKKUR VA INNOVATSION G'UYALAR



KLASTERLASH USULLARI YORDAMIDA AXBOROTNI IZLASH SAMARADORLIGINI OSHIRISH

Soyibova Bonu O'tkir qizi

*Shahrisabz Davlat Pedagogika Institutu Pedagogika Fakulteti
Talim tarbiya nazariyasi va metodikasi yo'nalishi Maktabgacha ta'lim Magistranti
[e-mail;utkirraupov62@gmail.com](mailto:utkirraupov62@gmail.com)*

Fan o'qituvchisi; Qodirov Farrux Ergash o'g'li

*Shahrisabz davlat pedagogika instituti "Matematika va ta'limda axborot
texnologiyasi" kafedrasini mudiri i.f.d DSc, dotsent.
<https://orcid.org/0000-0002-4574-7728>, [E-pochta:fgodirov@shdpi.uz](mailto:fgodirov@shdpi.uz)*

Annotatsiya *Mazkur maqolada klashterlash usuli mohiyati, ularning axborotni izlash jarayonidagi o'rni va samaradorligini oshirishdagi ahamiyati tahlil qilinadi. Shuningdek, mashinaviy o'rganish algoritmlari yordamida ma'lumotlarni guruhlash va qidiruv natijalarini optimallashtirish masalalari yoritiladi.*

Kalit so'zlar: *klasterlash, axborot izlash, mashinaviy o'rganish, K-means, ierarxik klasterlash, ma'lumotlar tahlili, algoritim, qidiruv tizimi.*

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ПОИСКА ИНФОРМАЦИИ С ПОМОЩЬЮ МЕТОДОВ КЛАСТЕРИЗАЦИИ

Соибова Бону Уткир кизи

*Шахрисабзский государственный педагогический институт, факультет
педагогики Направление: теория и методика обучения
Магистрант по направлению «Дошкольное образование»
E-mail: utkirraupov62@gmail.com*

Преподаватель предмета; Кодиров Фаррух Эргаши

*Заведующий кафедрой «Математика и информационные технологии в
образовании» Шахрисабзского государственного педагогического института,
доктор наук (DSc), доцент. ORCID: <https://orcid.org/0000-0002-4574-7728>
Электронная почта: fgodirov@shdpi.uz*

Аннотация: *В данной статье анализируется сущность метода кластеризации, его роль в процессе поиска информации и значение в повышении эффективности. Также рассматриваются вопросы группировки данных с использованием алгоритмов машинного обучения и оптимизации результатов поиска.*

Ключевые слова: *кластеризация, поиск информации, машинное обучение, K-means, иерархическая кластеризация, анализ данных, алгоритм, поисковая система.*



TANQIDIY NAZAR, TAHLILIIY TAFAKKUR VA INNOVATSION G‘OYALAR



IMPROVING THE EFFICIENCY OF INFORMATION RETRIEVAL USING CLUSTERING METHODS

Soyibova Bonu O‘tkir qizi

*Shakhrisabz State Pedagogical Institute, Faculty of Pedagogy Field: Theory and
Methodology of Education Master’s Student in Preschool Education*

E-mail: utkirraupov62@gmail.com

Subject teacher; Qodirov Farrux Ergash o‘g‘li

*Head of the Department of “Mathematics and Information Technologies in
Education” at Shakhrisabz State Pedagogical Institute, Doctor of Science (DSc),
Associate Professor. ORCID: <https://orcid.org/0000-0002-4574-7728>*

Email: fgodirov@shdpi.uz

Abstract: *This article analyzes the essence of the clustering method, its role in the information retrieval process, and its importance in improving efficiency. It also discusses issues of data grouping using machine learning algorithms and optimization of search results.*

Keywords: *clustering, information retrieval, machine learning, K-means, hierarchical clustering, data analysis, algorithm, search system.*

Zamonaviy axborot tizimlarida ma’lumotlar hajmining keskin ortishi ularni samarali izlash va qayta ishlash muammosini dolzarb qilib qo‘ymoqda. An’anaviy qidiruv tizimlari ko‘pincha katta hajmdagi ma’lumotlar ichidan aniq natijalarni ajratishda qiyinchilikka duch keladi [1, 30]. Shu sababli, klasterlash usullaridan foydalanish orqali ma’lumotlarni guruhlash va qidiruv natijalarini tartibga solish muhim ahamiyat kasb etadi. Mazkur maqolaning maqsadi — klasterlash usullari yordamida axborotni izlash samaradorligini oshirish imkoniyatlarini tahlil qilishdir. Klasterlash — bu o‘xshash ma’lumotlarni bir guruhga jamlash jarayoni bo‘lib, u ma’lumotlarni tahlil qilish va qidiruv jarayonini soddalashtirishga xizmat qiladi [2, 50].

Axborotni izlash tizimlarida klasterlash usullari qidiruv natijalarini guruhlash orqali foydalanuvchiga qulaylik yaratadi. Bu usul natijalarni mavzular bo‘yicha ajratib, kerakli ma’lumotni tez topishga yordam beradi [3, 65].

K-means algoritmi klasterlashning eng keng tarqalgan usullaridan biri bo‘lib, ma’lumotlarni oldindan belgilangan klasterlarga ajratadi. Bu algoritm tezkorligi va soddaligi bilan ajralib turadi [4, 70]. Ierarxik klasterlash esa ma’lumotlarni Klasterlash usullarining samaradorligini oshirishda masofa o‘lchovlari (distance metrics) muhim rol o‘ynaydi. Eng ko‘p qo‘llaniladigan usullar — Evklid masofasi, Manhattan masofasi va kosinus o‘xshashlik ko‘rsatkichidir. Ayniqsa, matnli ma’lumotlar bilan ishlashda kosinus o‘xshashlik yuqori aniqlik beradi [2, 95]. Axborotni izlash tizimlarida klasterlash jarayonida vektor fazo modeli (Vector Space Model) keng qo‘llaniladi. Bu modelda hujjatlar vektor ko‘rinishida tasvirlanib, ular o‘rtasidagi o‘xshashlik matematik usullar orqali aniqlanadi [1, 85].





TF-IDF (Term Frequency – Inverse Document Frequency) usuli matnli ma’lumotlarni tahlil qilishda muhim ahamiyatga ega. Bu usul so‘zlarning muhimlik darajasini aniqlab, klasterlash natijalarini yanada aniqroq qiladi [3, 110].

Klasterlashda optimal klasterlar sonini aniqlash ham muhim muammo hisoblanadi. Bu borada “Elbow method” va “Silhouette score” kabi usullar keng qo‘llaniladi [4, 100].

Shuningdek, klasterlash usullarida shovqinli (noisy) ma’lumotlar bilan ishlash muhim masala hisoblanadi. DBSCAN kabi algoritmlar shovqinli ma’lumotlarni aniqlash va ularni ajratib olish imkonini beradi [10, 140].

Klasterlash usullari axborotni izlash tizimlarida indekslash jarayonini optimallashtirishga yordam beradi. Ma’lumotlar oldindan klasterlarga ajratilgani sababli, qidiruv jarayoni tezlashadi va resurslardan samarali foydalaniladi [6, 115].

Dinamik klasterlash texnologiyalari real vaqt rejimida ma’lumotlarni qayta guruhlash imkonini beradi. Bu ayniqsa tez o‘zgaruvchan axborot tizimlarida muhim hisoblanadi [7, 120].

Klasterlash usullari yordamida tematik qidiruv tizimlari yaratish mumkin. Bu tizimlar ma’lumotlarni mavzular bo‘yicha ajratib, foydalanuvchiga aniq yo‘naltirilgan natijalarni taqdim etadi [8, 130]. Bundan tashqari, klasterlash usullari rekomendatsiya tizimlarida ham qo‘llaniladi. Masalan, foydalanuvchilarni qiziqishlariga qarab guruhlash orqali ularga mos mahsulot yoki xizmatlar tavsiya etiladi [9, 140]. Parallel hisoblash (parallel computing) texnologiyalari bilan birgalikda klasterlash usullari katta hajmdagi ma’lumotlarni tezkor qayta ishlash imkonini beradi. Bu esa zamonaviy axborot tizimlarining samaradorligini sezilarli darajada oshiradi [6, 125]. Klasterlash usullarining yana bir muhim yo‘nalishi — gibriddan foydalanishdir. Ya’ni, bir nechta algoritmlarni birlashtirib, yanada aniq natijalarga erishish mumkin. Masalan, K-means va ierarxik klasterlashni birgalikda qo‘llash yuqori samaradorlik beradi [5, 135]. Sun’iy intellekt asosida adaptiv klasterlash tizimlari foydalanuvchi xatti-harakatlariga mos ravishda o‘zini o‘zgartirib boradi. Bu esa qidiruv natijalarining doimiy ravishda yaxshilanishini ta’minlaydi [7, 130].

Shuningdek, klasterlash usullari axborot ortiqchaligini kamaytirishga xizmat qiladi. Keraksiz yoki takroriy ma’lumotlar chiqarib tashlanib, foydalanuvchiga faqat muhim natijalar taqdim etiladi [8, 140]. Umuman olganda, klasterlash usullari zamonaviy axborot tizimlarida nafaqat qidiruv jarayonini, balki ma’lumotlarni boshqarish va tahlil qilishni ham sezilarli darajada takomillashtiradi. daraxtsimon tuzilma asosida guruhlaydi. Bu usul ma’lumotlar o‘rtasidagi bog‘liqlikni chuqurroq tahlil qilish imkonini beradi [5, 80]. Klasterlash usullari qidiruv samaradorligini oshirishda muhim rol o‘ynaydi. Ular katta hajmdagi ma’lumotlarni qismlarga ajratib, qidiruv jarayonini tezlashtiradi [6, 90]. Klasterlash usullarini takomillashtirishda o‘lchamlarni kamaytirish (dimensionality reduction) muhim ahamiyatga ega. Yuqori o‘lchamli ma’lumotlar bilan ishlash murakkab bo‘lgani sababli, PCA (Principal Component Analysis) va t-SNE kabi usullar qo‘llanilib, ma’lumotlar soddalashtiriladi va klasterlash aniqligi oshiriladi [2, 110].



TANQIDIY NAZAR, TAHLILIIY TAFAKKUR VA INNOVATSION G'UYALAR



Yarim nazoratli klasterlash (semi-supervised clustering) usullari ham zamonaviy yondashuvlardan biri hisoblanadi. Bu usulda qisman belgilangan ma'lumotlar yordamida tizim yanada aniqroq klasterlashni amalga oshiradi [7, 140]. Klasterlash jarayonida xususiyatlarni tanlash (feature selection) muhim bosqich hisoblanadi. Keraksiz yoki ortiqcha belgilarni olib tashlash orqali algoritm samaradorligi oshiriladi va hisoblash resurslari tejaladi [4, 115].

Matnli ma'lumotlarni klasterlash alohida yo'nalish bo'lib, unda hujjatlar o'xshashligi asosida guruhlanadi. Bu usul qidiruv tizimlarida yangiliklar, maqolalar va ilmiy hujjatlarni tartiblashda keng qo'llaniladi [3, 120]. Zamonaviy tizimlarda chuqur o'rganish (Deep Learning) asosida klasterlash usullari ham rivojlanmoqda. Autoencoder va neyron tarmoqlar yordamida ma'lumotlarning yashirin tuzilmasi aniqlanadi va yanada aniq klasterlar hosil qilinadi [5, 145].

Onlayn (incremental) klasterlash usullari real vaqt rejimida yangi ma'lumotlar kelib tushganda tizimni qayta o'rgatmasdan yangilash imkonini beradi. Bu katta hajmdagi dinamik ma'lumotlar bilan ishlashda juda muhim [6, 135].

Klasterlash usullari axborotni filtrlash va spamlarni aniqlashda ham qo'llaniladi. O'xshash xususiyatlarga ega zararli yoki keraksiz ma'lumotlar alohida klasterga ajratilib, foydalanuvchiga ko'rsatilmaydi [8, 150].

Ko'p darajali (multi-level) klasterlash tizimlari ma'lumotlarni bir necha bosqichda guruhlash imkonini beradi. Bu yondashuv katta va murakkab ma'lumotlar bazasida yuqori aniqlikni ta'minlaydi [9, 155]. Klasterlash usullarining samaradorligi baholash mezonlari orqali aniqlanadi. Silhouette koeffitsienti, Davies-Bouldin indeksi va Dunn indeksi kabi ko'rsatkichlar klasterlash sifatini aniqlashda keng qo'llaniladi [4, 120]. Gibrid axborot izlash tizimlarida klasterlash boshqa usullar bilan birgalikda ishlatiladi. Masalan, semantik qidiruv va klasterlashni birlashtirish orqali yanada aniq natijalarga erishish mumkin [1, 95]. Shuningdek, klasterlash usullari vizual qidiruv tizimlarida ham qo'llanilib, rasm va video ma'lumotlarni guruhlash orqali tezkor izlash imkonini yaratadi [3, 130]. Mobil va veb ilovalarda klasterlash texnologiyalari foydalanuvchiga mos kontentni tez topishga yordam beradi. Bu esa foydalanuvchi tajribasini sezilarli darajada yaxshilaydi [6, 140]. Umuman olganda, klasterlash usullari zamonaviy axborot izlash tizimlarining samaradorligini oshirishda asosiy texnologiyalardan biri bo'lib, ularning qo'llanish doirasi tobora kengayib bormoqda.

Mashinali o'rganish algoritmlari yordamida klasterlash tizimlari doimiy ravishda takomillashib boradi. Bu esa qidiruv natijalarining aniqligini oshiradi [7, 100]. Klasterlash usullari foydalanuvchi tajribasini yaxshilaydi, chunki ular natijalarni tartibli va tushunarli shaklda taqdim etadi [8, 110].

Klasterlash usullari katta ma'lumotlar (Big Data) bilan ishlashda samarali hisoblanadi. Bu usullar turli manbalardan olingan ma'lumotlarni tahlil qilib, ularni tartibga soladi [9, 120]. Bulutli texnologiyalar bilan integratsiya klasterlash usullarining qo'llanish doirasini



kengaytiradi. Bu orqali katta hajmdagi ma'lumotlarni tezkor qayta ishlash imkoniyati yaratiladi [6, 95].

1. Klasterlash usullari turli sohalarda qo'llaniladi:
2. qidiruv tizimlarida
3. tibbiyotda ma'lumotlarni tahlil qilishda
4. biznesda mijozlarni segmentatsiya qilishda

Klasterlash usullarining yana bir muhim jihati — ma'lumotlar o'rtasidagi yashirin bog'liqliklarni aniqlashdir. Bu orqali tizim foydalanuvchiga yanada aniq va mos natijalarni taqdim etadi [2, 85]. Vizualizatsiya klasterlash natijalarini tushunishni osonlashtiradi. Grafik va diagrammalar orqali ma'lumotlar guruhlarini ko'rish mumkin [8, 120]. Real vaqt rejimida klasterlash texnologiyasi zamonaviy tizimlarda keng qo'llanilib, foydalanuvchilarga tezkor natijalar taqdim etadi [6, 105].

Xulosa qilib aytganda, klasterlash usullari axborotni izlash samaradorligini oshirishda muhim vosita hisoblanadi. Ular ma'lumotlarni guruhlash, qidiruv natijalarini tartibga solish va foydalanuvchiga qulay shaklda taqdim etish orqali qidiruv jarayonini optimallashtiradi. Mashinali o'rganish va sun'iy intellekt texnologiyalari bilan uyg'unlashgan holda klasterlash usullari zamonaviy axborot tizimlarining ajralmas qismiga aylanmoqda. Xulosa qilib aytganda, klasterlash usullari zamonaviy axborot izlash tizimlarining samaradorligini oshirishda muhim va ajralmas texnologiyalardan biri hisoblanadi. Ular katta hajmdagi ma'lumotlarni mantiqiy guruhlariga ajratish orqali qidiruv jarayonini soddalashtiradi, tezlashtiradi va natijalarning aniqligini sezilarli darajada oshiradi.

Klasterlash usullarining afzalligi shundaki, ular ma'lumotlar o'rtasidagi yashirin bog'liqliklarni aniqlash, ortiqcha va takroriy axborotni kamaytirish hamda foydalanuvchiga tartiblangan va tushunarli natijalarni taqdim etish imkonini beradi. Ayniqsa, K-means, ierarxik klasterlash va DBSCAN kabi algoritmlar turli turdagi ma'lumotlar bilan ishlashda o'zining yuqori samaradorligini namoyon etmoqda.

Shuningdek, klasterlash usullari mashinaviy o'rganish, sun'iy intellekt va katta ma'lumotlar texnologiyalari bilan integratsiyalashgan holda yanada rivojlanib bormoqda. Bu esa nafaqat qidiruv tizimlarida, balki tibbiyot, ta'lim, biznes va elektron tijorat kabi ko'plab sohalarda ham samarali qo'llanilishiga imkon yaratmoqda.

Zamonaviy yondashuvlar, jumladan chuqur o'rganish, onlayn klasterlash va gibridd modellar klasterlash jarayonini yanada mukammallashtirib, real vaqt rejimida yuqori aniqlikdagi natijalarni taqdim etishga xizmat qilmoqda. Shu bilan birga, o'lchamlarni kamaytirish va xususiyatlarni tanlash kabi usullar hisoblash jarayonini optimallashtirib, tizim samaradorligini oshiradi.

Kelajak istiqbolida klasterlash usullari yanada aqlli va moslashuvchan tizimlarga aylanishi kutilmoqda. Ularning semantik qidiruv, bilimlar grafigi va IoT texnologiyalari bilan integratsiyasi global axborot tizimlarini yangi bosqichga olib chiqadi. Bu esa foydalanuvchilarga yanada aniq, tezkor va ishonchli axborot olish imkonini beradi. Umuman olganda, klasterlash usullari axborotni izlash jarayonini tubdan takomillashtirib, uni samarali,





tizimli va foydalanuvchiga yo‘naltirilgan holga keltiradi. Shu sababli, ushbu texnologiyalarni o‘rganish va amaliyotga joriy etish zamonaviy axborot texnologiyalarining eng dolzarb yo‘nalishlaridan biri hisoblanadi. Kelajakda ushbu usullar yanada rivojlanib, katta hajmdagi ma‘lumotlar bilan ishlashda yanada samarali yechimlar taklif etishi kutilmoqda.

Foydalanilgan adabiyotlar

- 1) Manning C. Introduction to Information Retrieval. 2008.
- 2) Tan P. Introduction to Data Mining. 2019.
- 3) Qodirov, F. "" Қашқадарё худуди аҳолисига хизмат кўрсатиш тармоқлари ва уларга таъсир этувчи омиллар"." О ‘zbekiston Qishloq Va Suv xo‘jaligi" Jurnal." О ‘zbekiston Qishloq Va Suv xo‘jaligi" Jurnal (2022).
- 4) Qodirov, F. "" OPTIMUM SOLUTIONS FOR THE DEVELOPMENT OF MEDICAL SERVICES IN PRIVATE CLINICS". MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI QARSHI FILIALI." (2022).
- 5) Qodirov, F. "" QR-KOD TEXNOLOGIYASI ASOSIDA ELEKTRON KUTUBXONA TIZIMINI DASTURIY VA APPARAT TAMINOTINI YARATISH". MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI QARSHI FILIALI." (2021).
- 6) Qodirov, F. E., O. D. Doniyorov, and H. Shokirov Sh. "Basic Concepts Of Information Security In Information Systems. Wide Threats And Their Consequences." КОНЦЕПЦИИ УСТОЙЧИВОГО РАЗВИТИЯ НАУКИ В СОВРЕМЕННЫХ УСЛОВИЯХ (2021): 153-155.
- 7) Bozorova, Irina Jumanazarovna, and Dilfuzaxon Mamasharipovna Karayeva. "Modern programming technologies and their role." интеллектуальный капитал ххi века. 2020.
- 8) Kodirov, F. E., and J. E. Nematov. "BASIC TECHNOLOGY AND SERVICE MANAGEMENT MULTISERVICE NETWORKS." Инновации в технологиях и образовании: сб. ст. участников XII Между (2019): 214.
- 9) Qodirov, F. E., et al. "PROBLEMS AND SOLUTIONS FOR EFFECTIVE PROTECTION AGAINST NETWORK ATTACKS." НАУКОЕМКИЕ ИССЛЕДОВАНИЯ КАК ОСНОВА ИННОВАЦИОННОГО РАЗВИТИЯ 93 (2019).
- 10) Qodirov, F. E., J. U. Abdirasulov, and J. E. Nematov. "FORMING GOVERNMENT AGENCY WEBSITES WITH WORDPRESS CONTENT MANAGEMENT SYSTEM." Инновации в технологиях и образовании: сб. ст. участников XII Между (2019): 219.