

## MATN MA'LUMOTLARIDAN KERAKLI MA'LUMOTLARNI OLISH

**Primova Mastura Hakim qizi**

[primovamastura@navoiy-uni.uz](mailto:primovamastura@navoiy-uni.uz)

ToshDO'TAU o'qituvchisi

**Annotatsiya.** Matnli ma'lumotlarni intellektual tahlil qilish – strukturlanmagan ma'lumotlarni tahlil qilish va qayta ishlashning eng muhim usullaridan biri bo'lib, dunyodagi ma'lumotlarining qariyb 80 foizini tashkil qiladi. Bugungi kunda ko'pchilik tashkilot va muassasalar ma'lumotlar omborlari va bulutli platformalarda katta hajmdagi ma'lumotlarni to'playdi, saqlaydi. Har soniyada bir nechta manbalardan turli xil turdagi yangi ma'lumotlar axborot tizimiga qabul qilinadi va ular eksponent ravishda o'sishda davom etadi. Bugungi kunda katta hajmdagi ma'tnli ma'lumotlarni qayta ishlash NLPning dolzarb vazifasi hisoblanadi. NLP da har bir jumla bilan bog'liq ma'noni ajratib olish, uni kompyuter tushunadigan shaklga aylantirish uchun bir nechta mashina o'rganish algoritmlaridan foydalaniladi. Ushbu maqolada text mining, text analytics va text analysis farqlari haqida keltirilgan.

**Abstract.** Text mining is one of the most important methods for analyzing and processing unstructured data, accounting for about 80% of the world's data. Today, most organizations and institutions collect and store large amounts of data in data warehouses and cloud platforms. Every second, new data of various types from several sources enters the information system, and it continues to grow exponentially. Today, processing large volumes of text data is a pressing NLP task. NLP uses several machine learning algorithms to extract the meaning associated with each sentence and convert it into a form that a computer can understand. This article discusses the differences between text mining, text analytics, and text analytics.

**Аннотация.** Интеллектуальный анализ текстовых данных — один из важнейших способов анализа и обработки неструктурированных данных, на долю которого приходится около 80% мировых данных. Сегодня большинство организаций и учреждений собирают и хранят большие объемы данных в хранилищах данных и облачных платформах. Каждую секунду в информационную систему поступают новые данные разного типа из нескольких источников, и они продолжают расти в геометрической прогрессии. Сегодня обработка больших объемов текстовых данных является актуальной задачей НЛП. НЛП использует несколько алгоритмов машинного обучения для извлечения значения, связанного с каждым предложением, и преобразования его в форму, понятную компьютеру. В этой статье обсуждаются различия между интеллектуальным анализом текста, текстовой аналитикой и анализом текста.

**Kalit soʻzlar:** *Text Mining, Text analytics, text analysis, NLP, Concordance, Word frequency va Collocation.*

Hozirgi kunga matnli maʼlumotlar bizni qamrob olgan. Uygʻonanimizdan koʻp oʻtmay, odatda matnli xabarlar, elektron pochta xabarlar va ijtimoiy media xabarlar koʻrinishidagi katta hajmdagi maʼlumotlarni koʻrishimiz mumkin. Koʻp tashkilot va muassasalar maʼlumotlar omborlari va bulutli platformalarda katta hajmdagi maʼlumotlarni toʻplaydi va saqlaydi. Shuningdek, tashkilotlarda koʻplab manbaalar orqali: *ilovalar, veb-sahifalar, ijtimoiy media, mijozlar sharhlari va qoʻngʻiroq transkriptlaridan* toʻplangan maʼlumotlar katta hajmdagi matnli maʼlumotlarni tashkil qiladi. Bunday katta hajmli maʼlumotlardan kerakli maʼlumotlarni topib olish, tahlil qilish va qayta ishlash juda murakkab vazifalardan biri hisoblanadi. Bunday matnli maʼlumotlardan *kerakli va dolzarb maʼlumotlarni* olish uchun **text mining** jarayonidan va **text analysis** dasturi yordamida amalga oshiriladi.

**Text mining** – matnli maʼlumotlarni intellektual tahlili boʻlib, strukturlanmagan katta hajmdagi matn maʼlumotlarni tahlil qilish orqali, yangi maʼlumotlarni olish jarayonidir. Unda matn maʼlumotlardagi *faktlar, tendensiyalar, qonuniyatlar, konsepsiyalar, kalit soʻzlar va boshqa elementlarni* aniqlashga yordam beradi. Tashkilotlarda strukturlanmagan maʼlumotlarni strukturlangan maʼlumotlarga aylantirib, matnli hujjatlarning katta toʻplamlarini tahlil qiladi. Bu jarayon **text analysis** deb ham ataladi.

*Text mining quyidagi vazifalarini qamrab oladi:* matnni tasniflash, matnni klasterlash, taksonomiya yaratish, hujjatlarni umumlashtirish, obyektlarni ajratib olish va hissiyotlarni (sentiment) tahlil qilish.

Text mining bir nechta metodologiyalardan foydalanadi, ulardan biri *tabiiy tilni qayta ishlashdir (NLP)*. **Tabiiy tilni qayta ishlash (NLP)** - bu tabiiy tildan foydalangan holda odamlar va kompyuterlar oʻrtasidagi oʻzaro aloqani taʼminlab beradi. Ular kompyuter fanlari, tilshunoslik, maʼlumotlar fanlari va sunʼiy intellektning bir boʻlimi hisoblanib, kompyuterlarga inson tilining yozma va ogʻzaki shakllarini tushunishga imkon beradi. Olimlarning fikriga koʻra, 2025-yilga kelib NLP dunyo bozorida 43 milliard dollargacha baholanadi deb bashorat qilinadi.

Hozirgi kunda NLP ga asoslangan koʻplab ilovalar mavjud: Siri, Alexa va Google Assistant kabi ovozi yordamchilar soʻrovlarni tushunish va javoblar tayyorlash, Grammarly jummlarning grammatik toʻgʻriligini tekshirish va Google Tarjimon ham NLP orqali amalga oshirilgan.

NLPda har bir jumla bilan bogʻliq maʼnoni ajratib olish, uni kompyuter tushunadigan shaklga aylantirish uchun bir nechta mashina oʻrganish algoritmlaridan foydalaniladi. Bu vazifalarni bajarishda ikkita asosiy usullar bor: *semantik tahlil va sintaktik tahlil*.

**Semantik tahlil**- kompyuterlarga jummlar kontekstini inson tushungandek, tushunish imkonini yaratadi, yaʼni inson tilini tushunish jarayonidir. *Semantik tahlil*

*semantikaga asoslanadi* - matn orqali berilgan ma'noga asoslanadi. Unda gapning matn elementlarini aniqlab, ularga grammatik va semantik rolini belgilaydi. Natijada, bir nechta so'zlarga ega bo'lgan so'zlarning ma'nosini aniqlash uchun atrofdegan matnning kontekstini tahlil qiladi.

**Sintaktik tahlil** - tilning grammatik qoidalarga qanchalik rioya qilishini aniqlashda ishlatiladi. U sintaksisga asoslanib, bu tilshunoslikning jumladagi so'zlarni grammatik ma'noga ega bo'lishi uchun tartibga solish qoidalari bilan shug'ullanadi.

*NLPda qo'llaniladigan ba'zi sintaktik usullar:*

- *Nutqni belgilash qismi (Part-of-speech tagging):* har bir so'z uchun nutq qismini aniqlash
- *Jumlaning uzilishi (Sentence breaking):* katta matn bo'lagidan jumla chegaralarini topish
- *Morfologik segmentatsiya (Morphological segmentation):* so'zlarni individual qismlarga bo'lish
- *So'zlarni segmentatsiyalash (Word segmentation):* uzluksiz matnning katta qismlarini kichikroq alohida bloklarga bo'lish
- *Lemmatizatsiya (Lemmatization):* qulay tahlil qilish uchun so'zning flektiv shakllarini birlik shaklga o'tkazish
- *O'zaklash (Stemming):* flektiv so'zlarni ildiz shakllariga kesish
- *Sintaktik tahlil:* Gapning grammatik tahlilini bajarish.

### **Text mining nima uchun muhim?**

Ba'zi tashkilotlar katta hajmdagi matnli ma'lumotlarni to'plash imkoniyatiga ega. Misol uchun, mijozlarning sharhlari, mahsulot sharhlari va ijtimoiy media xabarlar orqali to'plangan ma'lumotlar katta hajmli ma'lumotlarning bir qismi hisoblanadi. Ushbu ma'lumotlar tashkilotlarga yuqori daromad keltiradi va foydalanuvchilarning talablariga mos tarzda qulayliklar yaratib berishi mumkin. Bunday ma'lumotlarning 80% strukturlanmagan ma'lumotlar hisoblanadi. Shuningdek, tashkilotlarga bir qancha noqulayliklar vujudga keladi. *Text mining* orqali matnlarni tasniflash (klassifikatsiyalash) jarayonini avtomatlashtirishga yordam beradi. Tasniflash bir nechta atributlarga asoslangan bo'lishi mumkin: *mavzu (topic)*, *niyat (intent)*, *hissiyot (sentiment)* va *til (language)*.

Foydalanuvchiga taklif qilinayotgan dasturiy ta'minot haqida qanday munosabatda ekanligini bilish kerak. Foydalanuvchi sharhlarini qo'lda ko'rib chiqish mumkin, ammo minglab sharhlar bo'lsa, *zerikarli jarayon* va *ko'p vaqt* talab qiladi. Bunday muammoni bartaraf etish uchun text mining yordamida bajarish mumkin. Text mining katta va murakkab ma'lumotlar to'plamini tez va oson tahlil qilib, ulardan kerakli ma'lumotlarni olish imkonini beradi. Bunday holda, text mining mahsulotning umumiy hissiyotini aniqlash imkonini beradi. Sharhlarning ijobiy, salbiy yoki neytral ekanligini aniqlash jarayoni *hissiyotlarni tahlil qilish* yoki *fikrni aniqlash* deb ataladi. Text mining orqali foydalanuvchilarga nimani yoqtirishi,

yoqtirmasligini yoki dasturning keyingi yangilanishga nimani kiritilishini xohlashini aniqlash uchun ishlatilishi mumkin. Foydalanuvchilarning muayyan mahsulotlar yoki mavzular bilan bog'liq holda foydalanadigan kalit so'zlarni aniqlashda ishlatiladi.

Shuningdek, tashkilotlar matn ma'lumotlarini chuqurroq o'rganishi, tegishli biznes tushunchalarini aniqlash yoki matnlar ichida qidiruv tizimlarini, an'anaviy ilovalar bilan aniqlanmaydigan o'zaro aloqalarni aniqlash uchun text mining vositalaridan foydalanishlari mumkin.

*Text miningdan tashkilotlar foyda olishlari mumkin bo'lgan ba'zi bir aniq usullar:*

- Farmatsevtika sanoatida maxfiy ma'lumotlarni ochish orqali, dori vosiolarini yanada rivojlantirish;
- Ishlab chiqaruvchi tashkilotlar real vaqt rejimida mijozlarning fikr-mulohazalarini tahlil qilishlari va darhol e'tibor berishni talab qiladigan mahsulotdagi xatolar yoki kamchiliklarni aniqlashlari;
- Kompaniyalar tuzilgan ma'lumotlarni yaratishi, ularni ma'lumotlar bazalariga integratsiyalashi va undan tavsiflovchi yoki bashoratli tahlil kabi katta ma'lumotlar tahlilining har xil turlari uchun foydalanishi.

Muxtassar qilib aytganda, text mining tashkilotlarga ma'lumotlarni ishga tushirishga, foydalanuvchilarni xursand qiladigan va natijada daromadlilik oshiradigan ma'lumotlarga asoslangan qarorlar qabul qilishga yordam beradi.

### **Text mining, text analytics va text analysis**

**Text mining** va **text analysis** ko'p hollarda sinonim sifatida ishlatiladi. Biroq, *text analysis* ikkalasidan farq qiladi.

**Text analytics** foydalanuvchilarga strukturlangan va strukturlanmagan matn ma'lumotlaridan ma'lumot olish imkonini beruvchi text analysis yoki text mining dasturiy ta'minoti sifatida ishlatiladi. Text mining va text analytics bir xil muammoni hal qiladi, ya'ni “xom” matn ma'lumotlarini tahlil qiladi. Ammo ularning natijalari sezilarli darajada farqlanadi. *Text mining* matn ma'lumotlaridan **sifatli natijalar** deb hisoblanishi mumkin bo'lgan kerakli ma'lumotlarni chiqaradi. Boshqa tomondan, *text analytics* **miqdoriy natijalar** sifatida ko'rish mumkin bo'lgan katta hajmdagi matn ma'lumotlaridagi tendensiyalar va qonuniyatlarni aniqlashga qaratilgan.

Boshqacha qilib aytganda; *text analytics* katta hajmdagi matn ma'lumotlarini tahlil qilish orqali grafik va jadvallar kabi vizual hisobotlarni yaratish uchun mo'ljallangan. Shuningdek, *text mining* tahlilni osonlashtirish uchun strukturlanmagan ma'lumotlarni strukturlangan ma'lumotlarga aylantiradi.

**Text mining** - bu misollardan o'rganish va yangi ma'lumotlardan natijalarni bashorat qilish mumkin bo'lgan modellarni yaratish uchun statistika, tilshunoslik va mashinani o'rganishga tayanadigan ma'lumotlar data mining kichik to'plami. Text

analytics ma'lumotlarni vizuallashtirish uchun text mining modellari tomonidan olingan ma'lumotlardan foydalanadi.

### **Text mining usullari**

Matn ma'lumotlaridan kerakli ma'lumotlarni olish uchun ko'plab text mining usullaridan foydalaniladi. Quyida bir nechta eng mashhur usullar keltirilgan:

- **Konkordans (Concordance)**

*Concardas* so'z yoki so'zlar to'plamini kontekstni aniqlashda ishlatiladi. Ya'ni “so'z” inson tilida turli xil ma'nolarni anglatishi sababli, so'zlarni moslashtirish tahlili kontekstga qarab so'zning aniq ma'nosini tushunishga yordam beradi.

- **So'z chastotasi (Word frequency)**

*So'z chastotasi* strukturlanmagan matn ma'lumotlarida so'zning necha marta takrorlanganini aniqlashda ishlatiladi. Misol uchun, mijozlar sharhlarida "bugs", "errors", "failure" kabi so'zlarning mavjudligini tekshiradi va bu atamalar tez- tez uchrab turishi orqali mahsulotlarni yangilab boriladi.

- **Ibora (Collocation)**

*Ibora* - ko'pincha birga kelgan so'zlar ketma-ketligi. “Decision making”, “time-consuming” va “keep in touch” ba'zi misollarni ko'rish mumkin. Iborani aniqlash matn tafsilotlarini yaxshilashi va matnни yaxshiroq tahlil qilish natijalariga olib keladi.

### **Foydalanilgan adabiyotlar:**

1. Aggarwal, C. C., & Zhai, C. X. (2013). An introduction to text mining. In *Mining Text Data* (Vol. 9781461432234). [https://doi.org/10.1007/978-1-4614-3223-4\\_1](https://doi.org/10.1007/978-1-4614-3223-4_1)
2. Kao, A., & Poteet, S. (2005). Text mining and natural language processing: introduction for the special issue. *SIGKDD Explor Newsl*, 7(1).
3. Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: Facebook and Twitter perspectives. *Advances in Science, Technology and Engineering Systems*, 2(1). <https://doi.org/10.25046/aj020115>
4. Zanini, N., & Dhawan, V. (2015). Text Mining: An introduction to theory and some applications. *Research Matters: A Cambridge Assessment Publication*, 19.