



MATNLARINI TEMATIK MODELLASHTIRISH VA TASNIFLASH USULLARI

Elov Botir Boltayevich

texnika fanlari falsafa doktori, dotsent. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti.

E-pochta: elov@navoiy-uni.uz

ORCID: 0000-0001-5032-6648

Aloyev Narzillo Raxmatilloevich

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti tayanch doktoranti.

E-pochta: vip.alayev@gmail.com

ORCID: 0009-0009-4625-5539

Annotatsiya. *Tematik modellashtirish – katta hajmdagi matnli hujjatlardagi mavzularni o'rganishga qaratilgan nazoratsiz modellar asosida so'zlar va hujjatlarning vaznli ro'yxatini aniqlash usuli. Tematik modellashtirish – sun'iy intellektdan foydalanadigan tabiiy tilni qayta ishlash texnologiyasi bo'lib, turli tashkilotlarga ish tizimini boshqarish, mijozlarga uzluksiz xizmat ko'rsatish va turli jarayonlarda qulay sharoit yaratish uchun xizmat qiladi. Har bir sohada bir nechta harakatlanuvchi qismlar mavjud. Masalan, mijozlar so'rovlarini qayta ishlash masalasini ko'rib chiqamiz. Tashkilotda mijozlarga xizmat ko'rsatish bo'limlari har kuni minglab mijozlar so'rovlari bilan shug'ullanadi va bir nechta takrorlanuvchi vazifalarni bajarishga to'g'ri keladi. Ushbu takrorlanuvchi jarayonlarni NLP algortimlari orqali avtomatlashtirish mumkin. Mijozlarga xizmat ko'rsatish xizmati nafaqat hayot tendensiyalaridan ortda qolmaslikka urinmoqda, balki deyarli barcha jamoalar, jumladan, moliya, HR, buxgalteriya, ishlab chiqarish va marketing ham o'z vaqtlarini kundalik vazifalarga sarflashmoqda. Yuqorida keltirilgan barcha masalalarning yechimi sifatida sun'iy intellekt (Artificial Intelligence, AI) usullaridan foydalanish mumkin. Tematik modellashtirish (Topic modeling) – buni amalga oshira oladigan AI usullaridan biridir. Tematik modellashtirish usullari vositasida tashkilotdagi jarayonlarni avtomatlashtirish va mablag'larni tejash uchun sun'iy intellekt metodlaridan foydalanadigan holda tashkilot faoliyati samaradorligini oshitish mumkin. Ushbu maqolada o'zbek tili matnlarini tematik modellashtirish va tasniflash usullari keltiriladi.*

Kalit so'zlar: *Tematik modellashtirish, tematik modellar, mashinali o'qitish, Yashirin semantik tahlil (LSA), Yashirin Dirixle tahlili (LDA), matnlarni klasterlash, mavzular tasnifi, NLP algortimlari.*



Kirish

So'nggi yillarda katta hajmdagi strukturlanmagan ma'lumotlar miqdori ortib borayotganligi sababli, foydali va muhim ma'lumotlarni olish jarayon murakkablashdi. Biroq, mashinali o'qitish metodlari vositasida ma'lumotlarni intellektual tahlil qilish natijasida biz izlayotgan ma'lumotlarni olishimiz mumkin. Matnni intellektual tahlil qilish sohasidagi shunday usullardan biri **tematik modellashtirish (Topic Modelling, TM)** hisoblanadi. Tematik modellashtirish – matnli hujjatda mavjud mavzularni avtomatik tarzda aniqlash va matn korpusi tomonidan ko'rsatilgan yashirin naqsh (shablon, qolip)larni olish jarayonidir. Tematik modellashtirishda katta hajmdagi hujjatlar to'plamlari tahlil qilinib, ulardagi yashirin (latent) mavzular aniqlanadi. Tematik modellashtirish (TM) – matndagi o'xshash so'zlarning klasterlari yoki guruhlarini aniqlash uchun nazoratsiz **mashinali o'qitish (machine learning, ML)** usullaridan foydalanadigan statistik modellashtirishning bir turidir [1]. Matnni intellektual tahlil qilishning ushbu usuli *oldindan aniqlanmagan teglar* yoki *o'quv ma'lumotlarisiz* strukturlanmagan ma'lumotlarni tushunish uchun matndagi semantik strukturalaridan foydalanadi [2].

Statistikada va tabiiy tilni qayta ishlashda muhim vosita bo'lgan tematik modellashtirish hujjatlar to'plamida mavjud bo'lgan mavhum "mavzularni" aniqlash uchun mo'ljallangan statistik modelni o'z ichiga oladi. TM matndagi yashirin semantik tuzilmalarni aniqlashga imkon beruvchi matnni intellektual tahlil qilish usuli bo'lib xizmat qiladi. Tematik modellashtirishdan foydalangan holda, tadqiqotchilar o'rganilayotgan hujjatlarga kiritilgan asosiy mavzular va tushunchalar haqida tushunchaga ega bo'lishlari mumkin.

Tematik modellashtirish oddiy iboralar yoki lug'atga asoslangan kalit so'zlarni qidirish usullaridan foydalanadigan qoidalarga asoslangan matnni aniqlash yondashuvlaridan farq qiladi. TM usuli matnlarning katta klasterlarida so'zlar to'plamini ("mavzular") aniqlash va tahlil qilish uchun foydalaniladigan nazoratsiz yondashuv hisoblanadi. Mavzular **"korpusda birgalikda uchraydigan atamalarining takrorlanuvchi namunasi"** sifatida belgilanishi mumkin. Yaxshi tematik model natijasida - "sog'liq", "shifokor", "bemor", "kasalxona" so'zlari uchun - "sog'liqni saqlash" mavzusi, "fermer", "ekinlar", "bug'doy" so'zlari uchun - "qishloq xo'jaligi" mavzusi aniqlanishi lozim.

Tematik modellar *hujjatlarni klasterlash, matnli ma'lumotlarning katta bloklarini shakllantirish, strukturlanmagan matndan ma'lumot olish va xususiyatlarni aniqlash* NLP vazifalari uchun juda foydali vosita. Misol uchun, onlayn nashlar o'z foydalanuvchilar sonini oshirish uchun **article recommendation engines** tematik modellaridan foydalanmoqda. Turli mutaxassislar ishga yollash sohalari uchun tematik modellardan foydalanadilar. Bunda ish ta'riflarining yashirin xususiyatlarini ajratib olish va ularni to'g'ri nomzodlarga moslashtirishda TM usuli qo'llaniladi. Shuningdek, *elektron pochta xabarlari, mijozlar sharhlari va foydalanuvchilarning ijtimoiy media profillarining katta ma'lumotlar to'plamini tashkil qilish* maqsasida TM modellari ishlatiladi.

Tematik modellar – matnli hujjatlar to'plamini kiritish sifatida qabul qiladigan va so'zlar hamda hujjatlarning vaznli ro'yxatini natija sifatida taqdim etadigan, mavzular to'plamini, tuzilmalarni o'rganadigan nazoratsiz ML modellar hisoblanadi. Matnli hujjatlardan aniqlangan mavzular tahlilchiga matn mazmunini tushunishga yordam beradi yoki boshqa NLP ilovalari



uchun so'zlar va hujjatlarning mavzuga asoslangan tasvirlarini yaratish uchun ishlatilishi mumkin. Temarik modellarni aniqlash usullari ishlab chiqilgan so'ng *matnni tahlil qilish va qayta ishlashning* ko'plab ilovalari, jumladan, *tadqiqot matnini tahlil qilish* [3], *ma'lumot qidirish* [4], *tabiiy tilni qayta ishlash* [5] va mavzuni aniqlash [6] bilan mashhur matnni tahlil qilish va qayta ishlash vositasiga aylandi.

Tematik modellarning keng qo'llanilishiga qaramay, TM algoritmlarining stoxastik xususiyati tufayli tasodifiy o'zgarishlar va xatolar yuzaga kelishi mumkin. Ushbu muammoni bartaraf qilish uchun tematik modelni baholashning bir qancha usullari ishlab chiqilgan [7,8,9,10]. Ushbu usullar tematik modellarni tahlil qilish va sharhlanadigan mavzularga ega modellarni qurish uchun vositalar va ko'rsatkichlarni taqdim etishga qaratilgan. Masalan, modellar mavzuning muvofiqligi [10] yoki model barqarorligi o'lchovlari yordamida baholanishi mumkin. O'xshash mavzulardan izchil xulosa chiqarish xususiyati haqida [9] ishda fikr-mulohazalar keltirilgan.

Matn to'plamlaridagi mavzularni aniqlash uchun tematik modellashtirish usullaridan foydalanilganda, tabiiy ravishda savol tug'iladi: modeldan olingan mavzular tahlilchini qiziqtirgan mavzularga qanchalik mos keladi. Matnlarni tematik modellashtirish orqali *umumiy mavzularni aniqlash* va *adekvat klasterlashni ta'minlash* uchun hujjatlar tahlil qilinadi. Masalan, tematik modellashtirish algoritmi vositasida berilgan hujjatning *shartnoma*, *hisob-faktura*, *shikoyat* va boshqa turga mansub ekanligini ularning mazmuniga qarab aniqlashi mumkin.

Bugungi kunda **yashirin semantik tahlil (Latent semantic analysis, LSA)** va **yashirin Dirixle tahlili (latent Dirichlet analysis, LDA)** usullari orqali mavzularni toifalarga ajratish, zarur ma'lumotlarni taqdim etish va muhim qarorlar qabul qilishda yordam berish uchun katta hajmli matnli fayllarni tahlil qiladigan ikkita asosiy tematik modellashtirish yondashuvlaridan keng miqyosida foydalanilmoqda [2,7].

Yashirin semantik tahlil (LSA)

Yashirin semantik tahlil – bu matnning asosiy ma'nosini ajratib olish va ifodalashning statistik usuli hisoblanadi [11,12]. LSA usuli *ma'nosi o'xshash so'zlarning* kontekstda birgalikda ishlatilishi tamoyiliga asoslanadi. Masalan, LSA usuli foydalanuvchi bilan teskqri aloqani ta'minlash uchun foydalanuvchi profillarini avtomatik yaratishga imkon beradi. Tashkilot o'z navbatida o'xshash foydalanuvchilarni guruhlash va foydalanuvchi profillari to'plamini yaratish uchun *demografiya*, *xaridlar tarixi* va *oldingi o'zaro munosabatlar va xatti-harakatlardan* foydalanishi mumkin. Ushbu amallar bir martalik jarayon bo'lib, keyinchalik ML vositalari yordamida avtomatlashtirilishi mumkin.

LSA usuli so'zlarni kontekstdagi holati va hujjatdagi chastotasi bo'yicha semantik bog'laydi. Bunday holda, xatti-harakatlari va o'zaro munosabat turi kabi o'xshash atributlarga ega bo'lgan mijozlar bir xil mavzu (toifa) ostida guruhlanadi [11,13]. *Nazoratsiz mahsinali o'qitish (unsupervised machine learning)* usulidan foydalangan holda, LSA oldingi kirish va chiqish ma'lumotlari asosida avtomatik ravishda yangi mavzularni yaratadi. LSA, agar so'zlarning chastotasi va tartibi bir xil bo'lsa, barcha shunga o'xshash hujjatlar bir xil shablonlarga ega bo'ladi degan taxminni nazarda tutadi. Bu esa o'z navbatida birgina hujjatni emas balki hujjatlar to'plamini tahlil qilishga yordam beradi. Bu tabiiy tilni qayta ishlashning bir qismi bo'lib, inson tili va mulohazalarini o'rganish va tushunish bilan chambarchas bog'liq bo'lib, hujjatlarni va tasvirlarning tahlilini o'z ichiga oladigan ko'plab tadbirlari mavjud.



Yashirin Dirixle tahlili (LDA)

Yashirin Dirixle tahlili – tematik modellashtirishning eng mashhur usullaridan biri bo'lib hisoblanadi. Ushbu usul hujjatdagi so'zlar o'rtasidagi munosabatlar vositasida mavzular bo'yicha guruhlash orqali to'plamdagi yashirin strukturasi aniqlab beradi. U hujjatlarni mavzular "aralashma"si, mavzularni esa so'zlar "aralashma"si sifatida qabul qiladi. Bir nechta hujjatlarni tahlil qilganda, ularni turli xil taqsimotlarga ega o'xshash mavzular sifatida qaraydi. Masalan, A hujjatida "**Y mavzu: 15%**" va "**Z mavzu: 85%**"ni tashkil qilsin. B hujjatida esa "**Y mavzu: 30%**" va "**Z mavzu: 70%**"dan iborat bo'lishi mumkin. LDA modeli bir nechta mavzular umumiy so'zlarni o'z ichiga olishi mumkin degan taxmin bilan, mavzularni so'zlarga ajratadi.

Ushbu usulida hujjatlarni yozish jarayonining generativ ko'rinishiga asoslanadi. LDA algoritmi asosida "*Muallif Toshkentdagi shaxmat olimpiadasi haqida maqola yozishga qaror qiladi*" gapiga mos "**sport**" va "**O'zbekiston**" mavzularini aniqlaydi. Keyingi qadamda har bir so'z uchun ikkita mavzudan birini mos qo'yadi. LDA usuli faqat yakuniy hujjatlarni tahlil qiladi va asosiy mavzularni va ular qanday ko'rinishini taxmin qiladi. Bugungi kunda LDA usulining ko'plab kengaytmalari va moslashuvlari mavjud [2,7]:

- *Teglangan LDA (Labeled LDA)* [14];
- *Muallif-mavzu modeli (Author-topic model)* [15];
- *Dinamik mavzu modeli (Dynamic topic model)* [16];
- *Ierarxik mavzu modeli (Hierarchical topic model)* [17].

Bu modellar asosan ehtimollik nazariyasi va statistika usullariga tayanadi.

LDA usulida **hujjat-atama matritsasi (document term matrix, DTM)**ni shakllantirish orqali o'zaro bog'langan mavzular tarmog'ini ishlab chiqish va ularning *mazmuni, matni, tasvirlari* va boshqalarni *tasniflash* uchun bir nechta hujjatlarni tahlil qilish imkoniyati yaratiladi [18].

Tematik modellashtirishda mavzu o'xshash so'z namunalarini mavzularga guruhlash orqali **mavzular klasterlari** shakllantiriladi [19]. Zamonaviy kiyimlarni sotadigan chakana tijorat kompaniyasi misolida tematik modellashtirish masalasini ko'rib chiqamiz. Onlayn do'konda har bir toifadagi mahsulotlarning *turlari, rasmlari va tavsiflari* mavjud. Agar siz xaridorlarga o'xshash mahsulotlarni ko'rsatishni yoki mahsulotlarni tegishli toifalarga guruhlashni istasangiz, har bir mahsulotni o'z toifasiga qo'lda guruhlashni amalga oshirish unumdorlik jihatdan noto'g'ri amal hisoblanadi. Ishlarni osonlashtirish uchun tematik modellashtirish algoritmidan foydalangan holda, *elementlarni (tovarlar, narsalar)* avtomatik ravishda turli toifalarga ajratib guruhlash mumkin. Tematik model *so'z chastotasi, tartibi, masofasi* va *ma'nosiga* asoslangan shablon (qolip)larni topish uchun katta hajmdagi strukturalanmagan ma'lumotlarni tahlil qiladi va har xil elementlarni oldindan tayyorgarlikka ega bo'lmagan holda tegishli toifalarga guruhlaydi.

Tematik modellashtirish va klasterlash

Tematik modellashtirish – hujjatlar to'plamidagi yashirin mavzularni aniqlashning statistik usuli bo'lib hisoblanadi. **Klasterlash** – o'xshash ma'lumotlarni guruhlashga mo'ljallangan mashinali o'qitish usulidir [20]. Ikkala usul ham hujjatlarni guruhlaydi, lekin buni qanday tartibda amalga oshirish masalasida farq qilishadi. Klasterlash algoritmlari o'xshash elementlarni birlashtirib guruhlaydi, tematik modellashtirish algoritmlari esa elementlar orasidagi munosabatlarni aniqlaydi. Tematik modellashtirish hujjatlar to'plamida yashirin mavzularni topish



uchun statistik yondashuvdan foydalanadi. Klasterlash odatda elementlarni guruhlab bir butun tarzda tahlil qilish maqsadida ishlatiladi. Tematik modellashtirish elementlar orasidagi munosabatlarni topadi va ma'lumotlar to'plamining yashirin strukturasi tushunish imkonini beradi. Bundan tashqari, tematik modellashtirish ma'lumotlar nishon (metka)larini qo'lda belgilashni talab qilmaydi, klasterlash esa aksincha talab qiladi.

Tematik modellashtirish va mavzular tasnifi (topic classification)

Tematik modellashtirish va mavzularni tasniflash – ikkala usul ham nazoratsiz o'rganish ML usullaridir. Har ikkala usul ham o'rganish uchun belgilangan/teglangan ma'lumotlarni talab qilmaydi. Ularning ikkalasi ham ma'lumotlardagi yashirin shablon (pattern)larni aniqlash uchun ishlatilishi mumkin. Ushbu usullarning asosiy farqi ularning ushbu amalni turli yondashuv asosida amalga oshirishidir.

1-jadval. Tematik modellashtirish va mavzular tasnifi

Tematik modellashtirish	Mavzular tasnifi
Har qanday turdagi ma'lumotlar uchun umumiy nazoratsiz yondashuv qo'llaniladi	Bu hujjatlarni oldindan belgilangan mavzularga ajratadigan nazorat ostidagi yondashuv
Tematik modellashtirish hujjatlar to'plamidagi yashirin mavzularni aniqlaydi	Mavzular tasnifi hujjatga uning mavzusiga ko'ra belgi/teg qo'yadi

Tematik modellashtirish va matn tasnifi

Tematik modellashtirish hujjatlar to'plamidan mavzularni aniqlashni o'z ichiga olsa, matn tasnifi hujjatga uning kontenti asosida teglash uchun matn tasniflagichlaridan foydalanadi. Matn tasnifi hujjatlarni oldindan belgilangan toifalarga ajratadi [21,22]. Tematik modellashtirish va matn tasnifi o'rtasidagi bir nechta asosiy farqlar quyidagi jadvalda keltirilgan.

2-jadval. Tematik modellashtirish va matn tasnifi

Tematik modellashtirish	Matn tasnifi
Tematik modellashtirish nazoratsiz ML usuli bo`lib, ya'ni yorliqli/ teglangan o'quv ma'lumotlarini talab qilmaydi. Mashinali o'rganish algoritmlari shunchaki hujjatlarni ko'rib chiqadi va ularni mavzularga ajratadi.	Matnni tasniflashda matnni qanday tasniflashni o'rganish uchun nazorat ostidagi ML usuli va razmetkalangan/teglangan o'quv ma'lumotlaridan foydalaniladi.
Mavzuni modellashtirish usuli matn ma'lumotlarida yashirin tuzilmalarni aniqlashi mumkin.	Matnlarni tasniflash ko'proq matnlarga teg belgilashga qaratilgan.
Mavzuni modellashtirish matn tasnifiga usuliga qaraganda ancha moslashuvchan bo`lib, oldindan belgilangan toifalar to'plamini talab qilmaydi. Ushbu usul yordamchi vositalarsiz ko'p vaqt talab qiladigan va katta hisoblashlarni talab qiladi.	Matnni tasniflash ma'lumotlarning oldindan teglanganishini talab qiladi.



Tematik modellashtirish va matn tasnifi bir-biridan ishlash tamoyillari bilan farq qilsa-da, NLP vazifalarini hal qilishdan eng yaxshi AI ko'rsatkichlariga erishishi uchun ulardan birgalikda foydalanish mumkin. O'zbek tilidagi strukturalanmagan ma'lumotlar, jumladan, rasmlar, hujjatlar va matnlar bilan ishlaydigan sun'iy intellektni avtomatlashtirish vositasini yaratish bugungi kunda aktual NLP vazifalaridan hisoblanadi. Ushbu turdagi NLP ilovalari nazorat qilinadigan va yarim nazorat qilinadigan ML o'rganish usullarini birlashtiradi va sizning foydalanish imkoniyatlaringizga moslashtirilgan AI modellarini o'rgatish imkonini beradi.

Tematik modellashtirish ilovalari

Tematik modellashtirish turli xil muammolarni hal qilish uchun ishlatilishi mumkin:

- *Hujjatlarni tasniflash;*
- *Mijozlarni qo'llab-quvvatlash tizimida buyurtmalarni belgilash;*
- *Mijozlar arizalarini aniqlash va qayta ishlash;*
- *Mijozlarning fikr-mulohaza (sharh)larini tahlil qilish.*

Hujjatlarni tasniflash

Tematik modellashtirish – bu kontekstni tushunish va yangi hujjatlarni belgilash/teglash uchun tabiiy tilni qayta ishlashdan (NLP) foydalanadigan nazoratsiz ML usuli. Tematik modellashtirish usuli vositasida avtomatik ravishda har bir hujjat mos mavzular orqali teglanadi. Misol uchun, LDA usulidan odatda ilmiy maqolalarni tasniflash va kun.uz, daryo.uz va qalampir.uz kabi onlayn nashr maqolalarini tasniflashga yordam beradi. Bugungi kunda kompyuter va axborot texnologiyalarining ortib borayotgan rivojlanishi bilan ko'plab ilmiy maqolalar onlayn va oflayn tarzda nashr etilmoqda va yangi tadqiqot sohalari doimiy ravishda yaratilganligi sababli, foydalanuvchilar o'zlarining qiziqarli ilmiy maqolalarini topish va tasniflashda juda ko'p muammolarga duch kelishmoqda. S.Kim va J.Gil tomondan amalga oshirilgan ilmiy tadqiqotda hujjatlarini o'xshash mavzularga ega bo'lishi mumkin bo'lgan mazmunli sinfga birlashtirishi mumkin bo'lgan tadqiqot ishlarini tasniflash tizimini taklif qilingan [23]. Ular tomonidan taklif etilgan tizim har bir maqolaning tezislardan va mavzulardan LDA sxemasi bo'yicha kalit so'zlarni ajratib olgan. Keyinchalik, K-means klasterlash algoritmi vositasida har bir ilmiy ishning TF-IDF qiymatlariga asoslanib, barcha maqolalarni o'xshash mavzular bilan tadqiqot ishlariga tasniflash uchun qo'llanilgan.

Mijozlarni qo'llab-quvvatlash tizimida buyurtmalarni belgilash

Shuningdek, siz mijozlarni qo'llab-quvvatlash xizmatidagi arizalarni avtomatik teglash uchun tematik modellashtirishdan foydalanishingiz mumkin. Arizalarning belgilangan ma'lumotlar to'plamidan foydalanib, tematik modelini o'rgatishdan boshlash mumkin. Keyinchalik esa o'rganilgan mavzular o'z navbatida yangi arizalarni belgilash uchun ishlatilishi mumkin.

Machine Learning texnikasi mijozlarning yillar davomida duch kelgan muammolari haqida ma'lumotlarni to'playdi va sodir bo'layotgan asosiy muammolarni aniqlab beradi.

Bu qo'llab-quvvatlash xodimlariga muammolarni va ularning sodir bo'lish chastotasini tezda aniqlashga yordam beradi va mijozlarga eng yaxshi yordam ko'rsatishni ta'minlaydi hamda,



murakkab muammolarni tegishli bo'limga yo'naltiradi va shu orqali mijozlar tegishli ma'lumotlarni tezroq olishlari mumkin.

Mijozlar arizalarini aniqlash va qayta ishlash

Tematik modellashtirish usulidan mijozlarga xizmat ko'rsatish xizmatidagi qo'ng'iroqlarning dolzarbligini aniqlashda foydalanish mumkin. Mijoz so'rovlarini muhimlik darajasi bo'yicha tartiblash tashkilot xodimlari uchun haqiqiy muammo hisoblanadi. Tabiiy tilni qayta ishlash texnikasi va ML modellaridan foydalangan holda, tematik modellashtirish usuli orqali mijoz arizalardan mavzularni *ajratib olish*, ularni *guruhlash*, ular orasidagi *semantik munosabatlarni aniqlash*, arizalarni *turli toifalarga ajratish* va eng ko'p e'tibor talab qiladigan *arizalarni belgilash* imkonini beradi.

Mijozlarning fikr-mulohazalarini tahlil qilish

Ko'pgina kompaniyalar odatda qo'lda mijozlarning fikr-mulohazalarini tahlil qilishga e'tibor bermaydilar. Tematik modellashtirish yordamida mijozlarning sharhlarini osongina baholash va hatto mablag'ni tejash mumkin. Tematik modellashtirish yordamida mijozlarning mulohazalarini baholanadi va kommentariyalari asosida teglar ishlab chiqiladi. Ishlab chiqilgan teglardan foydalanish oson, tez va samaralidir.

Matn tasniflash ilovalari

Matnni tasniflash boshqariladigan ML algoritmlaridan foydalanadi. Quyida matnni tasniflashning ba'zi ilovalari keltirilgan:

- *So'rov natijalarini tahlil qilish (Analyze survey results);*
- *Murojaatlar marshrutini avtomatlashtirish (Automatic ticket routing);*
- *Hissiyot tahlili (Sentiment analysis);*
- *Elektron pochta xabarlarini tasniflash (Categorize email);*
- *Ijtimoiy tarmoqlarni "tinglash" (Social listening).*

So'rov natijalarini tahlil qilish

Matn tasnifi usuli vositasida so'rov natijalarini tahlil qilish maqsadida matn tasniflagichini o'rgatish orqali yangi natijalarni tasniflashga yordam beradi. Tashkilot biznesini rivojlantirish uchun mijozlar va xodimlarning siz haqingizda qanday fikrda ekanligini aniqlash va tahlil qilish muhim. Ularning fikr-mulohazalarini qo'lda yig'ish zerikarli bo'lib, so'rovnomalar yordamida ushbu vazifani hal qilish mumkin. *Net Promoter Score (NPS)* kabi mashhur so'rovnomalar xizmati juda muhim bo'lib, ammo so'rov matnlarini qo'lda tahlil qilish ko'p vaqt talab etadi. Ushbu jarayonni avtomatlashtirish orqali topshiriqni bajarish uchun soatlab vaqt sarflash zaruratini bartaraf etadi.

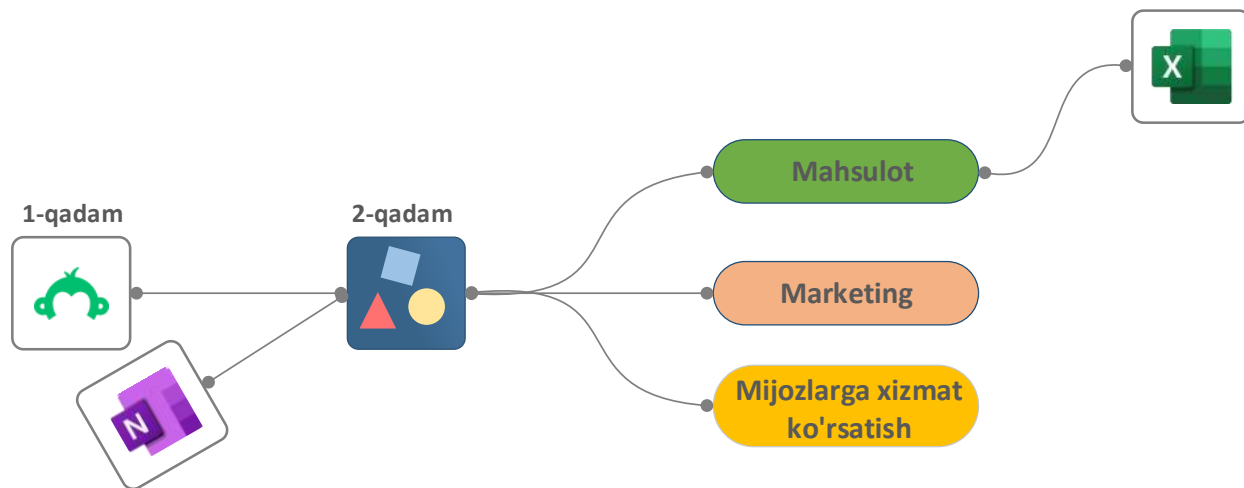
Jarayonni soddalashtirish uchun *matnlarni tasniflash ilovalari* mijozlar sharhlarini avtomatik ravishda *qo'llab-quvvatlovchilar*, *qoralovchilar* va *himoyachilar* kabi turli toifalarga bo'lib chiqadi. Shuningdek, u asosiy *shablonlarni aniqlash* va *fikr-mulohazalarning maqsadi va mavzusini tushunish* uchun **ijobiy**, **salbiy** va **neytral** javoblarni avtomatik ravishda aniqlaydi. Hissiyotlarni tahlil NLP ilovasi orqali mijozlarning fikr-mulohazalarini tahlil qilish va fikr-mulohazalarni avtomatik ravishda tegishli jamoa yoki rahbarlarga yo'naltirish imkonini beradi. Ular esa o'z navbatida xodim yoki mijozlar bilan ishlash tajribasini yaxshilash uchun tegishli choralarni ko'rishlari mumkin bo'ladi. Hissiyotlarni tahlil NLP ilovasi yordamida ushbu jarayonni



avtomatlashtirish orqali kompaniyadagi haftalar, hatto oylar qo'l mehnatini tejash mumkin. Ushbu NLP ilovasi yordamida quyida keltirilgan bir nechta tahlillarni amalga oshirish mumkin:

- *Ijobiy, salbiy va neytral javoblarni avtomatik aniqlash;*
- *Fikr-mulohazaning maqsadi va mavzusini tushunish;*
- *Filtrlashni tezda amalga oshirish.*

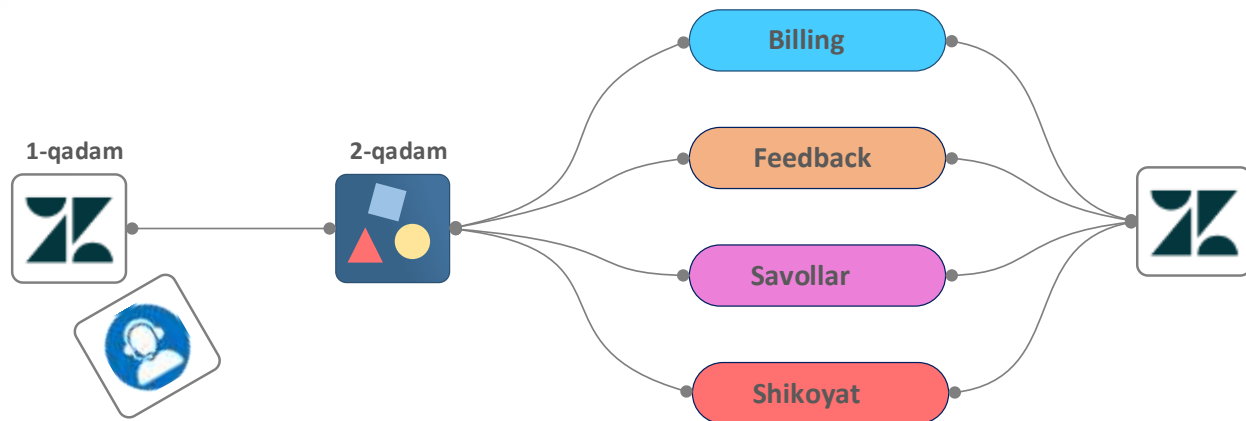
Quyidagi 1-rasmda mijozning bilan teskari aloqani tahlil qilishni *matnlarni tasniflash ilovasi* orqali amalga oshirilish misoli keltirilgan:



1-rasm. Matnlarni tasniflash jarayoni

Murojaatlar marshrutini avtomatlashtirish

Murojaatlar marshrutlarini mijozlarga xizmat ko'rsatish xodimiga avtomatik yo'naltirishda matn tasnifi foydali vosita hisoblanadi. Murojaatlarni qo'lda yo'naltirish uchun qo'shimcha xodim kerak bo'ladi. Bu esa tashkilot qo'llab-quvvatlash xizmati bo'limidagi xodimlarning vaqtini behuda sarflanishiga olib keladi. Chunki ko'pchilik chiptalar takroriy ahamiyatga ega. ML algoritmlari ushbu jarayonni avtomatlashtirish va xatolarni kamaytirish uchun maxsus nazorat qilinadi va o'qitiladi (2-rasm). Murojaatlarni avtomatik teglash va marshrutlash qo'llab-quvvatlash xizmati jarayonini sezilarli darajada yaxshilashi va muammolarni tezda hal qilish orqali mijozlar murojaatining yechimini yaxshilashi mumkin. Arizalarni avtomatik yo'naltirish haqida gap ketganda, *murojaatlar marshrutini avtomatlashtirish ilovasi* mijozlarga xizmat ko'rsatishni samarali tashkil qilishni ta'minlaydigan eng yaxshi yechimlardan biridir. Murojaatlarni (arizalarni) tasniflashni avtomatlashtirish orqali xodimlarning vaqtini tejagan holda ko'proq muhim mijozlarga ko'proq vaqt ajratishiga, ma'naviy holatlariga, ish unumdorligiga va mijozlarga tez orada javob qaytarishiga vaqt sarflashni ta'minlaydi.

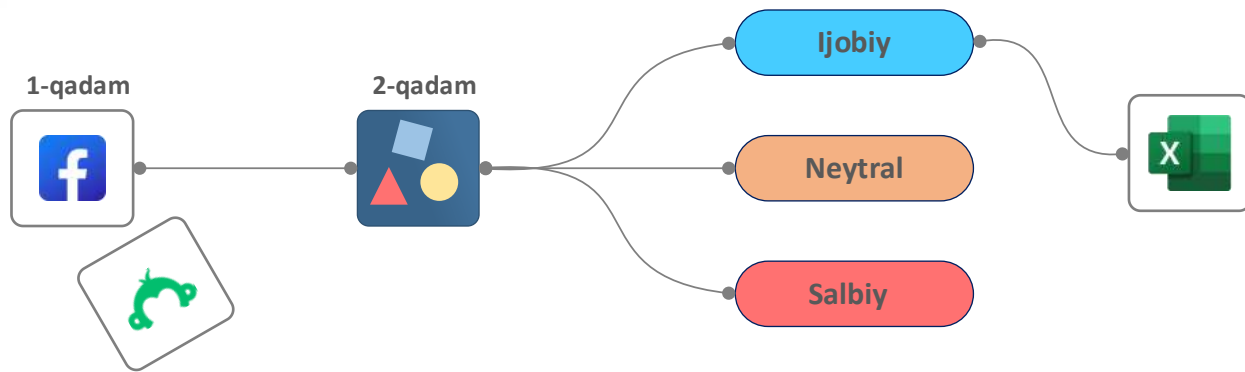


2-rasm. Murojaatlar marshrutini avtomatlashtirish jarayoni

Hissiyot tahlili

So'rovnomma yoki *teskari aloqa (feedback)* asosida hissiyotni tushunish va uni *ijobiy, salbiy* yoki *neytral* deb tasniflash o'z navbatida mijozga xizmat ko'rsatish muhitida juda muhimdir. Ushbu amal orqali mijozlarning ehtiyojlarini o'z vaqtida qondirilishiga yordam beradi. Arizadagi xabarda qanday hissiyot o'rin topganini bilish g'azablangan mijozlarning shikoyatlarini ijobiy hal qilishga yordam beradi. *Hissiyot tahlili ilovasi* orqali har qanday xabarni ochish yoki o'qishdan oldin uning qaysi hissiyot turiga tegishli ekanligini ajratib oladi va xabarlarini o'sha hissiyot asosida tasniflaydi (3-rasm). Hissiyot tahlili ilovasi quyidagilarga imkon beradi:

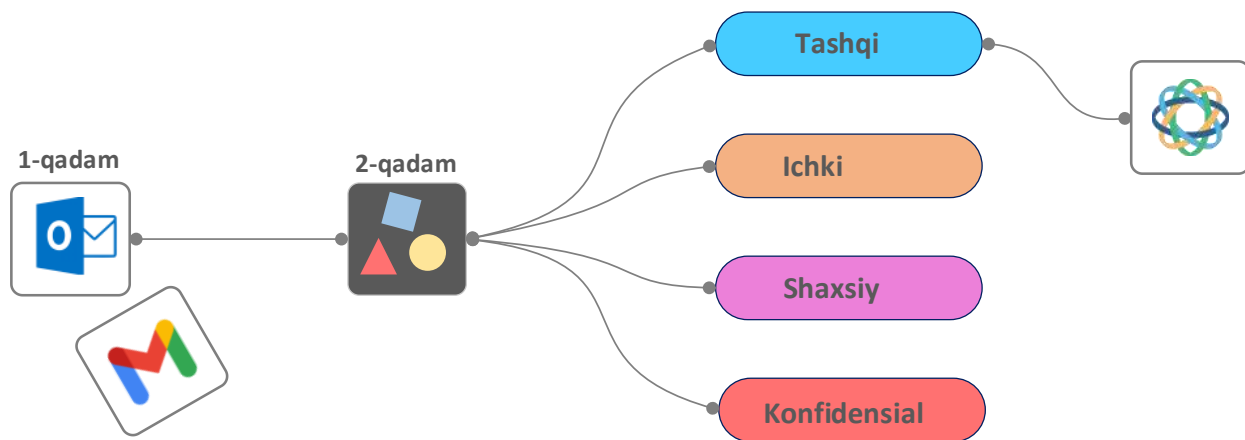
- **Mijozlarni qo'llab-quvvatlash murojaatlariga ustunlik berish:** mijozlarning salbiy fikr-mulohazalarini birinchi o'ringa qo'yis yoki hatto uni avtomatik ravishda nazoratchi darajasiga ko'tarish.
- **Mijozlarning qoniqishini yaxshilash:** vaqt o'tishi bilan fikr-mulohaza tendentsiyalarini kuzatib borish, avtomatlashtirilgan xabarlarini yuborish yoki Net Promoter Score (NPS) so'rovlaridagi ochiq savollarni belgilash.
- **Mijozlarni segmentatsiyasini yaxshilash:** Muayyan segmentni tushunish uchun his-tuyg'u tahlilini miqdoriy ma'lumotlar bilan bog'lash. Masalan, veb-saytda kamroq vaqt sarflaydigan mijozlar ko'proq pessimistik yoki yo'qligini aniqlash. Bu ularning veb-saytga pul sarflashiga xalaqit berishi mumkinmi, degan xulosaga kelish.
- **Odamlarning muayyan harakatlarga qanday munosabatda bo'lishini kuzatib borish:** odamlarning ma'lum bir o'zgarish haqida qanday fikrda ekanligi haqida tushunchaga ega bo'lish yoki kelajakdagi voqealarni kutish.
- **Tashkilot brendi monitoringini amalga oshirish:** Internetdagi tashkilot brendi eslatmalarga (ayniqsa salbiy sharhlarga) tezda munosabat bildirish.



3-rasm. Hissiyot tahlili jarayoni

Elektron pochta xabarlarini tasniflash

Elektron pochta xabarlarini boshqarish axborot texnologiyalari sohasidagi ko'pchilik mutaxassislar uchun asosiy muammolardan biri hisoblanadi. Elektron xabarlarini kontent asosida qo'lda toifalarga ajratish va teglash murakkab jarayon hisoblanadi. Elektron xabarlarini o'chirish, saqlash va muhimlik darajasiga ko'ra ajratish ko'p vaqt talab qilishi mumkin. Ba'zi hollarda, muhim xabarlarini o'tkazib yuborishimiz ehtimoldan holi emas. Sun'iy intellektga asoslangan matn tasnifi yuqorida keltirilgan holatlar uchun foydali bo'lishi mumkin. Elektron pochta xabarlarini tasniflash pochta xabarlarini shaxsiy, mijoz, ichki va tashqi kabi turli toifalarga ajratishga yordam beradi.



4-rasm. Elektron pochta xabarlarini tasniflash jarayoni

Elektron pochta xabarlarini tasniflash ilovalarining AI mexanizmi elektron pochta xabarlarini turli pochta qutilari (Gmail, Outlook va boshqalar) bo'ylab tasniflashni osonlashtiradi. AI modeli elektron pochta xabarlarini tasniflaganidan so'ng, NLP ilovasi keyingi bosqichlarni avtomatlashtirish imkonini beradi. Ushbu bosqichda elektron pochta xabarlarini xodimlarning ma'lum bir a'zosiga yo'naltirish yoki ularni ma'lum bir papkaga ko'chirishni o'z ichiga oladi.

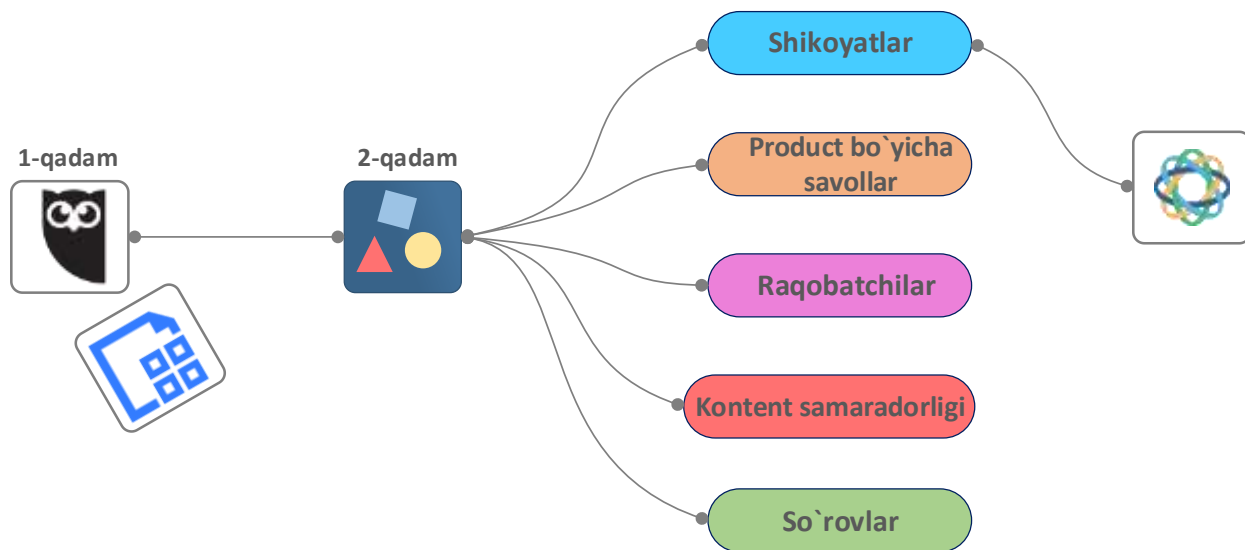
Ijtimoiy tarmoqlarni "tinglash"

Kompaniyalar o'z faoliyatlarini asosan Internet texnologiyalari orqali y'olga qo'rganlari sababli, asosiy jarayonlarni onlayn tarzda amalga oshirmoqdalar. Shu sababli kompaniya brendini yuksaltirish uchun onlayn tarmoqda qanday o'rin tutayotganini kuzatishi kerak. Ijtimoiy tarmoqlardagi brend imidjini saqlab qolish, o'z navbatida mijozlar muammolarini hal qilish uchun ham asosiy turtki bo'la oladi. Biroq, kompaniya haqida aytilgan hamma narsani qo'lda bajarish oson ish emas. Shuning uchun tashkilot biznes-jarayonlarini avtomatlashtirish kerak bo'ladi.



Mavjud va potentsial mijozlar fikr-mulohazalarini har doim tinglash va matn tasnifi orqali tashkilot brendini rivojlantirish uchun turli NLP usullaridan foydalanish lozim.

Ijtimoiy tarmoqlarni “tinglash” NLP ilovalari mijozlarning tashkilot brendi haqida nima fikr bildirishayotganlarini yaxshiroq tushunish uchun shikoyatlardan tortib raqobatchilargacha bo'lgan kommentariyalarni onlayn kuzatish imkonini beradi. Ushbu turdagi NLP ilovalari turli xil manbalardan tashkilot brendi uchragan kommentariyalarni va sharhlarni olish, ularni AI matn tasniflash modeli yordamida qayta ishlash va keyingi qadamlarni avtomatlashtirish imkonini beradi.



5-rasm. Ijtimoiy tarmoqlarni “tinglash” jarayoni

Mavzularni tasniflash

Mavzular tasnifi – hujjatlarni oldindan belgilangan mavzularga ajratishga asoslangan bo'lib, dastlab qiyin vazifadek tuyulishi mumkin. Ammo kerakli NLP ilova bilan buni bir tugmani bosish bilan amalga oshirishingiz mumkin. *Mavzularni tasniflash NLP ilovasi* – bu avtomatlashtirilgan, kodsiz AI yechimi bo'lib, u mavzularni oson va aniq toifalarga ajrata oladi. Mavzularni tasniflash NLP ilovasini quyidagi ketma-ketlikda shakllantirish mumkin:

- *Ma'lumotlarni yig'ish/to'plash;*
- *Tematik modelni o'rgatish;*
- *Yangi hujjatlarni teglash.*

Ma'lumotlarni yig'ish

Mavzuni tasniflashda *Mavzularni tasniflash NLP ilovasidan* foydalanish uchun avval belgilangan hujjatlarning ma'lumotlar to'plami kerak bo'ladi. Ko'p hollarda ma'lumotlar to'plami *matn faylida* yoki *.CSV* formatdagi faylda bo'lishi kerak. Ushbu fayllarning har bir satr hujjatni ifodalaydi. Hujjatlar misol sifatida *"sog'liq"*, *"moliya"* yoki *"ta'lim"* kabi toifalar bo'yicha teglanishi mumkin.

Tematik modelni o'rgatish

Keyingi qadamda teglangan ma'lumotlar to'plami asosida AI blokini o'rgatish lozim. AI bloklar o'rgatilganida so'ng keyingi qadamda ishlash samaradorligini yaxshilash lozim. Bunday holda, AI bloklar modelning ishlashini baholab berishi va uni yaxshilash uchun bajarilishi kerak



bo'lgan harakatlarni ko'rsatadi. Masalan, ko'rsatkichlarni yaxshilash uchun ma'lum miqdordagi teglarni yaratish yoki ko'proq ma'lumotlarni qo'shish mumkin. Ushbu bosqichlarni amalga oshirgandan so'ng, optimal natijaga erishib, AI blokini qayta o'qitish mumkin.

Yangi hujjatlarni teglash

Yuqoridagi qadamlar bajarilishi natijasida AI bloki asosida yangi hujjatlarni teglash mumkin. Modelni o'rgatish va turli testlardan o'tkazgandan so'ng, u kelajakdagi barcha hujjatlar bilan qanday ishlash kerakligini aniq biladi.

Matnlarini tematik modellashtirish va tasniflashda AI usullarida foydalanish orqali tashkilotga muntazam tarzda eng samarali vazifalarni tavsiya qilish imkoniyatini taqdim etadi. Matnlarini tematik modellashtirish va tasniflash NLP ilovasi foydalanuvchilarning odatlarini va shablonlarini o'rganadi. Shuningdek, u batafsil fikr-mulohazalarni bildirish uchun tashkilot faoliyati samaradorligini oshirish haqidagi ma'lumotlarni to'playdi. Kodsiz AI yechimi tashkilotlarga oddiy vazifalarni avtomatlashtirish va boshqa muhimroq vazifalarga e'tibor qaratish imkoniyatini taqdim etadi.

Xulosa

NLPda mavzularni modellashtirish – bu katta hajmdagi matnlarni avtomatik ravishda umumlashtirish uchun ishlatilishi mumkin bo'lgan algoritmlar to'plami. Matn xususiyatlari soni juda ko'p bo'lgani sababli modellarni o'qitish qiyinlashadi va modellarning samaradorligi pasaytiradi. YDA usuli NLPda tematik modellashtirishning taqsimlangan usuli bo'lib, hujjatlardan yashirin mavzularni avtomatik ravishda aniqlay oladi. Klasterlash – bu xususiyatlar yoki kuzatishlar bo'yicha guruhlashni aniqlashning yana bir nazoratsiz usuli bo'lib, tematik modellashtirishni va o'lchovni qisqartirishni turli xil algoritmlar bilan amalga oshirilishi mumkin: Non-negative Matrix Factorization, LSA, PLSA, Lda2Vec, Bert Topic-BERT. Ushbu maqolada tematik modellashtirishning tavsifi, qo'llanish sohalari, usullari, yondashuvlari haqida ma'lumot keltirildi. Shuningdek, tematik modellashtirish, mavzular tasnifi, klasterlash va matn tasnifi kabi NLP vazifalarining o'xhash va farqli jihatlari tushuntirildi. Tematik modellashtirishning zamonaviy ilovalari va ularni amalga qo'llanilishi tavsiflandi va tashkilot faoliyati samaradorligini oshirishga xizmat qiladigan amallari tavsiflandi. Tematik modellar hujjatlarni klasterlash, matnli ma'lumotlarning katta bloklarini shakllantirish, strukturlanmagan matndan ma'lumot olish va xususiyatlarni aniqlash NLP vazifalari uchun juda foydali vosita bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar

1. Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11). <https://doi.org/10.1007/s11042-018-6894-4>
2. Korencic, D., Ristov, S., Repar, J., & Snajder, J. (2021). A Topic Coverage Approach to Evaluation of Topic Models. *IEEE Access*, 9. <https://doi.org/10.1109/ACCESS.2021.3109425>
3. Chuang, J., Ramage, D., Manning, C. D., & Heer, J. (2012). Interpretation and trust: Designing model-driven visualizations for text analysis. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/2207676.2207738>



4. Wei, X., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006*. <https://doi.org/10.1145/1148170.1148204>
5. Li, L., Roth, B., & Sporleder, C. (2010). Topic models for word sense disambiguation and token-based idiom detection. *Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2010-July*.
6. Evans, M. S. (2014). A computational approach to qualitative analysis in large textual datasets. *PLoS ONE, 9*(2).
<https://doi.org/10.1371/journal.pone.0087908>
7. <https://doi.org/10.1371/journal.pone.0087908>
8. Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *ACM International Conference Proceeding Series, 382*.
<https://doi.org/10.1145/1553374.1553515>
9. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*.
10. Waal, A. de, & Barnard, E. (2008). Evaluating topic models with stability. *Annual Symposium of the Pattern Recognition Association of South Africa*.
11. Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*.
12. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6). [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASII>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9)
13. Tao, R., Wei, Y., & Yang, T. (2021). Metaphor Analysis Method Based on Latent Semantic Analysis. *Journal of Donghua University (English Edition), 38*(1).
<https://doi.org/10.19884/j.1672-5220.202010087>
14. Qi, Q., Hessen, D. J., Deoskar, T., & van der Heijden, P. G. M. (2023). A comparison of latent semantic analysis and correspondence analysis of document-term matrices. *Natural Language Engineering, 8*(10). <https://doi.org/10.1017/S1351324923000244>
15. Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*.
16. Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., & Steyvers, M. (2010). Learning author-topic models from text corpora. *ACM Transactions on Information Systems, 28*(1).
<https://doi.org/10.1145/1658377.1658381>
17. <https://doi.org/10.1145/1658377.1658381>
18. Gao, Q., Huang, X., Dong, K., Liang, Z., & Wu, J. (2022). Semantic-enhanced topic evolution analysis: a combination of the dynamic topic model and word2vec. *Scientometrics, 127*(3).
<https://doi.org/10.1007/s11192-022-04275-z>



19. Zou, X., Zhu, Y., Feng, J., Lu, J., & Li, X. (2019). A novel hierarchical topic model for horizontal topic expansion with observed label information. *IEEE Access*, 7. <https://doi.org/10.1109/ACCESS.2019.2960468>
20. B.Elov, Z.Xusainova, N.Xudayberganov. O`zbek tili korpusi matnlari uchun TF-IDF statistik ko`rsatkichni hisoblash. *SCIENCE AND INNOVATION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8 UIF-2022: 8.2 | ISSN: 2181-3337*
21. Elov B., Aloyev N., Yuldashev A. SVD va NMF metodlari orqali tematik modellashtirish // *Труды XI Международной конференции «Компьютерная обработка тюркских языков» «TURKLANG 2023»*. Бухара, 20-22 октября 2023 г.
22. Lad, R. Y., & Metkewar, P. S. (2020). Bibliometric survey on clustering algorithm of machine learning and its application. *International Journal of Scientific and Technology Research*, 9(4).
23. Lavanya, P. M., & Sasikala, E. (2021). Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network: A comprehensive survey. *2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021*.
<https://doi.org/10.1109/ICSPC51351.2021.9451752>
24. <https://doi.org/10.1109/ICSPC51351.2021.9451752>
25. Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>
26. Kim, S. W., & Gil, J. M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-Centric Computing and Information Sciences*, 9(1). <https://doi.org/10.1186/s13673-019-0192-7>