

Multilingual Embedding modellərinin Azərbaycan dilində performansı

Başlıq

3 əsas dil modelindən istifadə olundu:

- XLM-RoBERTa;
- BGE-M3;
- LaBSE

Bu modellərin sayını artırmaq da olar. Bu modellərin hər biri BERT tipli modellərdir. BERT modellərinin müxtəlif versiyaları. BERT modellərini istifadə etməkdə əsas məqsəd odur ki, onlar auto-regressive decoder arxitekturasında yox, bi-directional encoder-decoder arxitekturasında olduğuna görə embedding-ləri yalnız özündən əvvəlki kontekstə əsasən yox, həm də özündən sonrakı kontekstə əsasən yaradır.

Əsas hissə

Prosesin ümumi work flow-u belədir:

- 1) PDF dokumenti oxunur;
- 2) PDF-dəki mətn təmizlənir;
- 3) Mətn chunk-lara ayrılır. Hər bir chunk-un maksimum uzunluğu 1024 xarakterdir.
- 4) Hər bir chunk dokumenti üçün 3 fərqli vektor embedding-ləri alınır. Hər bir vektor embedding-ləri fərqli vector database-də yaddaşa saxlanılır. Ümumilikdə 3 database var: XLM-RoBERTa modelinin yaratdığı vektor embedding-ləri, BGE-M3 modelinin yaratdığı vektor embedding-ləri və LaBSE modelinin yaratdığı vektor embeddin-lərinin saxlanıldığı database.
- 5) Sorğu yazılır. Daha sonra bu sorğunun 3 fərqli vektor embedding-i çıxarılır.
- 6) 3 fərqli vektor embedding-dən istifadə etməklə vector database-də axtarış və retrieval prosesi aparılır.

P.S. Qeyd edim ki, istifadə edilən vector database-i chromadb alətidir. Və sadəcə raw retrieval-in nəticələri görmək istədiyimizə görə chunking və retrieval prosesləri xüsusi arxitektura və texnika ilə yox, sadə məntiqlə hazırlanıb. Searching isə cosine similarity-ə əsasən aparılıb.

P.S. Retrieval prosesində ən uyğun olan 3 və ya 5 dokument retrieve edilir.

Sorğu

Nəzarətsiz öyrənmə nədir?

Dokumentə əsasən düzgün cavab belədir:

4.2. Nəzarətsiz Öyrənmə (Unsupervised Learning)

Nəzarətsiz öyrənmədə, alqoritm yalnız giriş məlumatları əsasında öyrənir və bu məlumatlardakı struktur və əlaqələri aşkar etməyə çalışır. Bu növ öyrənmə, məlumatların klasterləşdirilməsi və ya gizli strukturların aşkarlanması üçün faydalıdır.

Əsas nəzarətsiz öyrənmə alqoritmlərinə aşağıdakılar daxildir:

- **K-Means:** Məlumatları K sayda klasterə bölür və hər bir nöqtəni ən yaxın klaster mərkəzinə aid edir.
- **Hierarchical Clustering:** Məlumatları iyerarxik struktura malik klasterlərə bölür.
- **Principal Component Analysis (PCA):** Məlumatların ölçülərini azaldaraq əsas komponentləri saxlayır.
- **Association Rule Learning:** Məlumatlar arasındakı əlaqə qaydalarını aşkar edir, məsələn, birlikdə alınan məhsulları təyin etmək üçün.

- **XLM-RoBERTa**

Ən yaxın 3 chunk dokumenti:

- 1) Kompüter Görmə Modelləri və Alqoritmləri Kompüter görmə sahəsində istifadə olunan əsas modellər və alqoritmlər aşağıdakılardır: • Konvolyusional Sinir Şəbəkələri (CNN): Görüntülərin təsnifatı və obyekt tanıma üçün ən geniş istifadə olunan dərin öyrənmə modeli. • YOLO (You Only Look Once): Real vaxt rejimində obyekt tanıma üçün istifadə olunan model. • R-CNN (Region-based CNN): Görüntüdəki obyektlərin yerini təyin etmək və tanımaq üçün istifadə olunan model. • U-Net: Tibbi görüntülərdə segmentasiya üçün istifadə olunan model. • OpenPose: İnsan bədən quruluşunu təyin etmək üçün istifadə olunan model. • Face Recognition: Üz tanıma üçün istifadə olunan modellər və alqoritmlər. 6.3. Azərbaycanda Kompüter Görmənin Tətbiqi Azərbaycanda

kompyuter görmə texnologiyaları müxtəlif sahələrdə tətbiq olunur: • Təhlükəsizlik: Videomüşahidə sistemlərində üz tanıma, şübhəli davranışların aşkarlanması. • Nəqliyyat: Avtomobil nömrələrinin tanınması, trafik monitorinqi, sürət pozuntularının aşkarlanması.

- 2) • Kənd təsərrüfatı: Bitkilərin vəziyyətinin monitorinqi, xəstəliklərin aşkarlanması, məhsul yığımının avtomatlaşdırılması. • Tibb: Tibbi görüntülərin (MRT, KT) təhlili, xəstəliklərin diaqnostikası. • Pərakəndə satış: Rəflərin monitorinqi, müştəri davranışının təhlili. • İstehsalat: Keyfiyyətə nəzarət, istehsal proseslərinin avtomatlaşdırılması. Azərbaycanda kompyuter görmə sahəsində həyata keçirilən layihələrə aşağıdakılar daxildir: • "Təhlükəsiz şəhər": Bakı və digər şəhərlərdə videomüşahidə sistemlərinin tətbiqi və üz tanıma texnologiyalarından istifadə. • "Ağıllı nəqliyyat": Trafik axınını optimallaşdırmaq üçün videoanalizdən istifadə. • "Rəqəmsal kənd təsərrüfatı": Dronlardan istifadə edərək əkin sahələrinin monitorinqi və xəstəliklərin aşkarlanması. • Tibbi diaqnostika sistemləri: Tibbi görüntülərin avtomatik təhlili üçün sistemlərin yaradılması. 7.
- 3) Azərbaycan universitetlərində məlumat elmi üzrə təhsil proqramları təqdim edilir və bu sahədə mütəxəssislərin hazırlanması prosesi davam edir. Bundan əlavə, bir sıra özəl təlim mərkəzləri də məlumat elmi və maşın öyrənməsi üzrə kurslar təşkil edirlər. Azərbaycanda məlumat elminin əsas tətbiq sahələri aşağıdakılardır: • Maliyyə və bank sektoru: Kredit risklərinin qiymətləndirilməsi, fırıldaqçılığın aşkarlanması, müştəri segmentasiyası. • Telekommunikasiya: Şəbəkə idarəetməsi, müştəri davranışının təhlili, xidmət keyfiyyətinin proqnozlaşdırılması. • Neft və qaz sənayesi: Hasilat proseslərinin optimallaşdırılması, avadanlıqların profilaktik texniki xidməti, riskin qiymətləndirilməsi. • E-ticarət: Məhsul tövsiyələri, müştəri davranışının təhlili, tələbin proqnozlaşdırılması. • Dövlət sektoru: Vətəndaşlara xidmətlərin təkmilləşdirilməsi, resurların bölüşdürülməsinin optimallaşdırılması, dövlət siyasətinin qiymətləndirilməsi. 4.

• **BGE-M3**

Ən yaxın 3 chunk:

- 1) • Dəstək Vektor Maşınları (Support Vector Machines): Məlumatları müxtəlif kateqoriyalara ayıran optimal hiperplanlər tapır. • K-ən Yaxın Qonşu (K-Nearest Neighbors): Yeni nümunələri ən yaxın qonşularının xüsusiyyətlərinə əsasən təsnif edir. 4.2. Nəzarətsiz Öyrənmə (Unsupervised Learning) Nəzarətsiz öyrənmədə, alqoritm yalnız giriş məlumatları əsasında öyrənir və bu məlumatlardakı struktur və əlaqələri aşkar etməyə çalışır. Bu növ öyrənmə, məlumatların klasterləşdirilməsi və ya gizli strukturların aşkarlanması üçün

faydalıdır. Əsas nəzarətsiz öyrənmə alqoritmlərinə aşağıdakılar daxildir: • K-Means: Məlumatları K sayda klasterə bölür və hər bir nöqtəni ən yaxın klaster mərkəzinə aid edir. • Hierarchical Clustering: Məlumatları iyerarxik struktura malik klasterlərə bölür. • Principal Component Analysis (PCA): Məlumatların ölçülərini azaldaraq əsas komponentləri saxlayır. • Association Rule Learning: Məlumatlar arasındakı əlaqə qaydalarını aşkar edir, məsələn, birlikdə alınan məhsulları təyin etmək üçün. 4.3.

- 2) Maşın Öyrənməsi Alqoritmləri və Modelləri Maşın öyrənməsi alqoritmləri və modelləri, kompüterlərin məlumatlar əsasında öyrənmək və proqnozlar vermək qabiliyyətini təmin edir. Bu alqoritmlər müxtəlif kateqoriyalara bölünür: 4.1. Nəzarətli Öyrənmə (Supervised Learning) Nəzarətli öyrənmədə, alqoritm giriş məlumatları və onlara uyğun çıxış nəticələri əsasında öyrənir. Məqsəd, yeni giriş məlumatları üçün dəqiq çıxış nəticələrini proqnozlaşdırma bilən bir model yaratmaqdır. Əsas nəzarətli öyrənmə alqoritmlərinə aşağıdakılar daxildir: • Xətti Reqressiya (Linear Regression): Bir və ya bir neçə müstəqil dəyişən əsasında asılı dəyişəni proqnozlaşdırmaq üçün istifadə olunur. • Loqistik Reqressiya (Logistic Regression): İkili təsnifat problemləri üçün istifadə olunur, məsələn, e-poçtun spam olub-olmadığını təyin etmək. • Qərar Ağacları (Decision Trees): Məlumatları bir sıra qərar qaydalarına əsaslanaraq təsnif edir. • Təsadüfi Meşə (Random Forest): Bir çox qərar ağacının nəticələrini birləşdirərək daha dəqiq proqnozlar verir.
- 3) Gücləndirilmiş Öyrənmə (Reinforcement Learning) Gücləndirilmiş öyrənmədə, agent ətraf mühitlə qarşılıqlı əlaqədə olaraq, aldığı mükafat və ya cəzalara əsasən optimal strategiyanı öyrənir. Bu növ öyrənmə, oyunlar, robot idarəetməsi və avtomatik ticarət kimi sahələrdə istifadə olunur. Əsas gücləndirilmiş öyrənmə alqoritmlərinə aşağıdakılar daxildir: • Q-Learning: Agent müxtəlif vəziyyətlərdə fərqli hərəkətlərin dəyərini öyrənir. • Deep Q Network (DQN): Q-learning alqoritmni dərin sinir şəbəkələri ilə birləşdirir. • Policy Gradient Methods: Agentin davranış siyasətini birbaşa optimallaşdırır. 4.4. Dərin Öyrənmə (Deep Learning) Dərin öyrənmə, çoxqatlı sinir şəbəkələrindən istifadə edərək böyük həcmli və kompleks məlumatları emal etmək üçün istifadə olunan maşın öyrənməsinin bir alt sahəsidir. Əsas dərin öyrənmə modellərinə aşağıdakılar daxildir: • Konvolyusional Sinir Şəbəkələri (Convolutional Neural Networks, CNN): Təsvirlərin emalı və tanınması üçün istifadə olunur.

- **LaBSE**

Ən yaxın 3 chunk dokumenti aşağıdakılardır:

- 1) Təhsil Sahəsində Süni İntellekt Təhsil sahəsində süni intellekt texnologiyaları tədris və öyrənmə proseslərini təkmilləşdirmək üçün istifadə olunur. Azərbaycanca təhsil sahəsində süni intellektin tətbiqinə aşağıdakı nümunələri göstərmək olar: • Fərdiləşdirilmiş təhsil: Tələbələrin fərdi ehtiyaclarına uyğun təhsil proqramlarının yaradılması. • Avtomatik qiymətləndirmə: İmtahanların və ev tapşırıqlarının avtomatik qiymətləndirilməsi. • Virtual müəllimlər və assistentlər: Tələbələrə 24/7 dəstək təqdim edən süni intellekt əsaslı sistemlər. • Təhsil analitikası: Tələbələrin uğurlarının və tədris prosesinin effektivliyinin təhlili. • Dil öyrənmə tətbiqləri: Azərbaycan dilini və xarici dilləri öyrənmək üçün süni intellekt əsaslı tətbiqlər. 7.3. Səhiyyə Sahəsində Süni İntellekt Səhiyyə sahəsində süni intellekt texnologiyaları diaqnostika, müalicə və xəstəliklərin profilaktikası kimi sahələrdə tətbiq olunur.
- 2) • Kibertəhlükəsizlik: Süni intellekt sistemlərinin kiber hücumlardan qorunması və təhlükəsizlik standartlarının müəyyən edilməsi. 8.3. Azərbaycanda Süni İntellektin Etik və Hüquqi Çərçivəsi Azərbaycanda süni intellektin etik və hüquqi aspektləri ilə bağlı müzakirələr davam edir və bu sahədə bir sıra təşəbbüslər həyata keçirilir: • Məlumatların qorunması üzrə qanunvericilik: Şəxsi məlumatların qorunması ilə bağlı qanunvericilik süni intellekt sistemlərinin istifadəsi kontekstində nəzərdən keçirilir. • E-dövlət sistemində süni intellektin tətbiqi: Dövlət xidmətlərində süni intellektin tətbiqi zamanı etik və hüquqi məsələlərin nəzərə alınması. • Təhsil və maarifləndirmə: Süni intellekt sahəsində etik və hüquqi məsələlər barədə təhsil və maarifləndirmə proqramlarının həyata keçirilməsi. • Beynəlxalq əməkdaşlıq: Süni intellektin etik və hüquqi aspektləri ilə bağlı beynəlxalq təşkilatlarla əməkdaşlıq və təcrübə mübadiləsi. 9.
- 3) Gücləndirilmiş Öyrənmə (Reinforcement Learning) Gücləndirilmiş öyrənmədə, agent ətraf mühitlə qarşılıqlı əlaqədə olaraq, aldığı mükafat və ya cəzalara əsasən optimal strategiyanı öyrənir. Bu növ öyrənmə, oyunlar, robot idarəetməsi və avtomatik ticarət kimi sahələrdə istifadə olunur. Əsas gücləndirilmiş öyrənmə alqoritmlərinə aşağıdakılar daxildir: • Q-Learning: Agent müxtəlif vəziyyətlərdə fərqli hərəkətlərin dəyərini öyrənir. • Deep Q Network (DQN): Q-learning alqoritmini dərin sinir şəbəkələri ilə birləşdirir. • Policy Gradient Methods: Agentin davranış siyasətini birbaşa optimallaşdırır. 4.4. Dərin Öyrənmə (Deep Learning) Dərin öyrənmə, çoxqatlı sinir şəbəkələrindən istifadə edərək böyük həcmli və kompleks məlumatları emal etmək üçün istifadə olunan maşın öyrənməsinin bir alt sahəsidir. Əsas dərin öyrənmə modellərinə aşağıdakılar daxildir: • Konvolyusional Sinir Şəbəkələri (Convolutional Neural Networks, CNN): Təsvirlərin emalı və tanınması üçün istifadə olunur.

Kompüter görmənin əsas komponentləri hansılardır?

Düzgün cavab:

6.1. Kompüter Görmənin Əsas Komponentləri

Kompüter görmə bir neçə əsas komponenti əhatə edir:

- **Görüntü əldə etmə:** Rəqəmsal kameralar və ya sensorlar vasitəsilə görüntülərin toplanması.
- **Görüntü emalı:** Görüntülərin keyfiyyətinin artırılması, küydən təmizlənməsi və xüsusiyyətlərinin çıxarılması.
- **Obyekt tanıma:** Görüntüdəki obyektlərin tanınması və təsnifatı.
- **Səhnə təhlili:** Görüntüdəki obyektlər arasındakı əlaqələrin və kontekstin başa düşülməsi.
- **Hərəkət təhlili:** Video görüntülərində obyektlərin hərəkətinin izlənməsi və təhlili.

• **XLM-RoBERTa**

Ən yaxın 3 dokument:

- 1) • Rekurrent Sinir Şəbəkələri (Recurrent Neural Networks, RNN): Ardıcıl məlumatların emalı üçün istifadə olunur, məsələn, mətn və ya zaman seriyası. • Long Short-Term Memory (LSTM): RNN-in bir növüdür və uzunmüddətli asılılıqları öyrənmək qabiliyyətinə malikdir. • Transformer: Diqqət mexanizmlərinə əsaslanan və təbii dil emalında geniş istifadə olunan model. • Generative Adversarial Networks (GAN): Yeni məlumatlar generasiya etmək üçün istifadə olunan iki sinir şəbəkəsinin qarşılıqlı rəqabətinə əsaslanan model. Azərbaycanda bu alqoritmlərin və modellərin tətbiqi müxtəlif sahələrdə həyata keçirilir. Məsələn, banklar kredit skoring sistemlərində nəzarətli öyrənmə metodlarından, pərakəndə satış şirkətləri müştəri segmentasiyası üçün nəzarətsiz öyrənmə metodlarından, oyun və əyləncə sənayesi isə gücləndirilmiş öyrənmə metodlarından istifadə edir. Dərin öyrənmə modelləri isə əsasən təsvir və mətn tanıma, təbii dil emalı və səs tanıma kimi sahələrdə tətbiq olunur. 5.
- 2) Məlumat elminin əsas komponentləri aşağıdakılardır: 1. Məlumatların toplanması: Müxtəlif mənbələrdən məlumatların əldə edilməsi və saxlanması. 2. Məlumatların təmizlənməsi və hazırlanması: Xam məlumatların analiz üçün hazırlanması, səhvlərin düzəldilməsi, çatışmayan dəyərlərin emal edilməsi. 3. Məlumat analizi: Statistik metodlar və alqoritmlər vasitəsilə məlumatların təhlili. 4. Maşın öyrənməsi modellərinin tətbiqi: Məlumatlar əsasında proqnozlaşdırma və təsnifat modellərinin yaradılması. 5. Nəticələrin vizuallaşdırılması və təqdimatı: Analiz

nəticələrinin anlaşılan formada təqdim edilməsi. Azərbaycanda məlumat elmi son illərdə sürətlə inkişaf etməkdədir. Bir çox şirkətlər və dövlət qurumları öz fəaliyyətlərində məlumat elmi metodlarını tətbiq etməyə başlamışlar. Məsələn, banklar kredit risklərinin qiymətləndirilməsində, telekommunikasiya şirkətləri müştəri davranışının təhlilində, neft şirkətləri isə hasilat proseslərinin optimallaşdırılmasında məlumat elmi üsullarından istifadə edirlər.

- 3) • Ağıllı şəbəkələr: Enerji şəbəkələrinin effektivliyinin artırılması və enerji itkisinin azaldılması. • Enerji səmərəliliyi: Binalarda və sənaye müəssisələrində enerji istehlakının optimallaşdırılması. • Alternativ enerji mənbələrinin idarə edilməsi: Günəş və külək enerjisi kimi alternativ enerji mənbələrinin effektiv idarə edilməsi.
8. Süni İntellektin Etik və Hüquqi Aspektləri Süni intellekt texnologiyalarının sürətli inkişafı bir sıra etik və hüquqi məsələləri gündəmə gətirir. Bu məsələlər həm global miqyasda, həm də Azərbaycan kontekstində müzakirə olunur.
- 8.1. Süni İntellektin Etik Məsələləri Süni intellektin etik məsələlərinə aşağıdakılar daxildir:
- Şəffaflıq və izah edilə bilən süni intellekt: Süni intellekt sistemlərinin qərar qəbul etmə prosesinin şəffaf və başa düşülən olması.
 - Ədalət və ayrı-seçkilik: Süni intellekt sistemlərinin ayrı-seçkiliyə yol verməməsi və ədalətli qərarlar qəbul etməsi.

• **BGE-M3**

Ən yaxın 3 chunk dokumenti:

- 1) • Tibb: Tibbi sənədlərin avtomatik emalı və təsnifatı. 6. Kompüter Görmə (Computer Vision) və Azərbaycanda Tətbiqi Kompüter görmə, kompüterlərin rəqəmsal görüntüləri və ya videoları başa düşmək və emal etmək qabiliyyətini təmin edən süni intellekt sahəsidir. Bu sahə, obyekt tanıma, üz tanıma, təsvir təsnifatı, hərəkət təhlili və s. kimi tətbiqləri əhatə edir.
- 6.1. Kompüter Görmənin Əsas Komponentləri Kompüter görmə bir neçə əsas komponenti əhatə edir:
- Görüntü əldə etmə: Rəqəmsal kameralar və ya sensorlar vasitəsilə görüntülərin toplanması.
 - Görüntü emalı: Görüntülərin keyfiyyətinin artırılması, küydən təmizlənməsi və xüsusiyyətlərinin çıxarılması.
 - Obyekt tanıma: Görüntüdəki obyektlərin tanınması və təsnifatı.
 - Səhnə təhlili: Görüntüdəki obyektlər arasındakı əlaqələrin və kontekstin başa düşülməsi.
 - Hərəkət təhlili: Video görüntülərində obyektlərin hərəkətinin izlənməsi və təhlili.
- 6.2.
- 2) Kompüter Görmə Modelləri və Alqoritmləri Kompüter görmə sahəsində istifadə olunan əsas modellər və alqoritmlər aşağıdakılardır:
- Konvolyusional Sinir Şəbəkələri (CNN): Görüntülərin təsnifatı və obyekt tanıma üçün ən geniş istifadə olunan dərin öyrənmə modeli.
 - YOLO (You Only Look Once): Real vaxt

rejimində obyekt tanıma üçün istifadə olunan model. • R-CNN (Region-based CNN): Görüntüdəki obyektlərin yerini təyin etmək və tanımaq üçün istifadə olunan model. • U-Net: Tibbi görüntülərdə segmentasiya üçün istifadə olunan model. • OpenPose: İnsan bədən quruluşunu təyin etmək üçün istifadə olunan model. • Face Recognition: Üz tanıma üçün istifadə olunan modellər və alqoritmlər. 6.3. Azərbaycanda Kompüter Görmənin Tətbiqi Azərbaycanda kompüter görmə texnologiyaları müxtəlif sahələrdə tətbiq olunur: • Təhlükəsizlik: Videomüşahidə sistemlərində üz tanıma, şübhəli davranışların aşkarlanması. • Nəqliyyat: Avtomobil nömrələrinin tanınması, trafik monitorinqi, sürət pozuntularının aşkarlanması.

- 3) • Kənd təsərrüfatı: Bitkilərin vəziyyətinin monitorinqi, xəstəliklərin aşkarlanması, məhsul yığımının avtomatlaşdırılması. • Tibb: Tibbi görüntülərin (MRT, KT) təhlili, xəstəliklərin diaqnostikası. • Pərakəndə satış: Rəflərin monitorinqi, müştəri davranışının təhlili. • İstehsalat: Keyfiyyətə nəzarət, istehsal proseslərinin avtomatlaşdırılması. Azərbaycanda kompüter görmə sahəsində həyata keçirilən layihələrə aşağıdakılar daxildir: • "Təhlükəsiz şəhər": Bakı və digər şəhərlərdə videomüşahidə sistemlərinin tətbiqi və üz tanıma texnologiyalarından istifadə. • "Ağıllı nəqliyyat": Trafik axınını optimallaşdırmaq üçün videoanalizdən istifadə. • "Rəqəmsal kənd təsərrüfatı": Dronlardan istifadə edərək əkin sahələrinin monitorinqi və xəstəliklərin aşkarlanması. • Tibbi diaqnostika sistemləri: Tibbi görüntülərin avtomatik təhlili üçün sistemlərin yaradılması. 7.

• **LaBSE**

Ən yaxın 3 dokument:

- 1) • Tibb: Tibbi sənədlərin avtomatik emalı və təsnifatı. 6. Kompüter Görmə (Computer Vision) və Azərbaycanda Tətbiqi Kompüter görmə, kompüterlərin rəqəmsal görüntüləri və ya videoları başa düşmək və emal etmək qabiliyyətini təmin edən süni intellekt sahəsidir. Bu sahə, obyekt tanıma, üz tanıma, təsvir təsnifatı, hərəkət təhlili və s. kimi tətbiqləri əhatə edir. 6.1. Kompüter Görmənin Əsas Komponentləri Kompüter görmə bir neçə əsas komponenti əhatə edir: • Görüntü əldə etmə: Rəqəmsal kameralar və ya sensorlar vasitəsilə görüntülərin toplanması. • Görüntü emalı: Görüntülərin keyfiyyətinin artırılması, küydən təmizlənməsi və xüsusiyyətlərinin çıxarılması. • Obyekt tanıma: Görüntüdəki obyektlərin tanınması və təsnifatı. • Səhnə təhlili: Görüntüdəki obyektlər arasındakı əlaqələrin və kontekstin başa düşülməsi. • Hərəkət təhlili: Video görüntülərində obyektlərin hərəkətinin izlənməsi və təhlili. 6.2.

- 2) Kompüter Görmə Modelləri və Alqoritmləri Kompüter görmə sahəsində istifadə olunan əsas modellər və alqoritmlər aşağıdakılardır: • Konvolýusional Sinir Şəbəkələri (CNN): Görüntülərin təsnifatı və obyekt tanıma üçün ən geniş istifadə olunan dərin öyrənmə modeli. • YOLO (You Only Look Once): Real vaxt rejimində obyekt tanıma üçün istifadə olunan model. • R-CNN (Region-based CNN): Görüntüdəki obyektlərin yerini təyin etmək və tanımaq üçün istifadə olunan model. • U-Net: Tibbi görüntülərdə seqmentasiya üçün istifadə olunan model. • OpenPose: İnsan bədən quruluşunu təyin etmək üçün istifadə olunan model. • Face Recognition: Üz tanıma üçün istifadə olunan modellər və alqoritmlər.
- 6.3. Azərbaycanda Kompüter Görmənin Tətbiqi Azərbaycanda kompüter görmə texnologiyaları müxtəlif sahələrdə tətbiq olunur: • Təhlükəsizlik: Videomüşahidə sistemlərində üz tanıma, şübhəli davranışların aşkarlanması. • Nəqliyyat: Avtomobil nömrələrinin tanınması, trafik monitorinqi, sürət pozuntularının aşkarlanması.
- 3) • Məxfilik və məlumatların qorunması: Şəxsi məlumatların süni intellekt sistemləri tərəfindən istifadəsi zamanı məxfiliyin qorunması. • Məsuliyyət və hesabatlılıq: Süni intellekt sistemlərinin qəbul etdiyi qərarlar üçün kimin məsuliyyət daşması. • İnsanın nəzarəti: Süni intellekt sistemlərinin insan nəzarəti altında olması və insanların həyatı və rifahı üçün təhlükə yaratmaması.
- 8.2. Süni İntellektin Hüquqi Aspektləri Süni intellektin hüquqi aspektlərinə aşağıdakılar daxildir: • İntellektual mülkiyyət hüquqları: Süni intellekt sistemləri tərəfindən yaradılmış məzmunun intellektual mülkiyyət hüquqları. • Məlumat qanunvericiliyi: Şəxsi məlumatların toplanması, saxlanması və istifadəsi ilə bağlı qanunvericilik. • Məsuliyyət hüququ: Süni intellekt sistemlərinin səbəb olduğu zərərlərin məsuliyyətinin müəyyən edilməsi. • Tənzimləmə və standartlar: Süni intellekt sistemlərinin yaradılması və istifadəsi üçün müəyyən standartların və tənzimləmə çərçivələrinin mövcudluğu.

Azərbaycan dilinin xüsusiyyətləri təbii dilin emalında hansı çətinliklər yaradır və bu çətinlikləri aradan qaldırmaq üçün hansı yanaşmalar mövcuddur?

Düzgün cavab:

Çətinliklər:

- **Morfoloji mürəkkəbliк:** Aqlütinativ dil olaraq, bir sözə çoxlu şəkilçi əlavə edilə bilər
- **Resursların məhdudluğu:** Annotasiya edilmiş korpuslar və lüğətlər azdır
- **Standartlaşdırılmış alətlərin azlığı:** Hazır TDE alətləri məhduddur

Həll yanaşmaları:

- **Xüsusi morfoloji təhlil sistemləri:** Azərbaycan dilinin şəkilçi sistemini nəzərə alan alqoritmlər
- **Transfer öyrənmə:** Digər türk dilləri üçün hazırlanmış modellərdən istifadə
- **Hibrid yanaşma:** Qayda əsaslı və statistik metodların kombinasiyası
- **AzerBERT kimi yerli dil modelləri:** Azərbaycan dili üçün spesifik transformerlər
- **Çoxdilli modellər:** XLM-RoBERTa kimi 100+ dildə öyrədilmiş modellərin adaptasiyası

- **XLM-RoBERTa**

'Gələcəkdə süni intellekt texnologiyalarının daha da inkişaf edəcəyi və cəmiyyətin müxtəlif aspektlərinə daha dərin təsir göstərəcəyi gözlənilir. Bu kontekstdə, texnologiyanın inkişafını etik və hüquqi çərçivədə idarə etmək, insan və süni intellekt əməkdaşlığının optimal formalarını tapmaq və texnologiyanın faydalarını ədalətli bölüşdürmək kimi məsələlər önəm kəsb edir. Azərbaycan da global süni intellekt inqilabının bir parçası olaraq, bu texnologiyaların inkişafında və tətbiqində öz yerini tutmağa çalışır. Bu səylər davam etdikcə, Azərbaycanın rəqəmsal transformasiyası və süni intellekt sahəsində potensialının reallaşdırılması üçün yeni imkanlar açılacaqdır.', 'Bu sahədə aparılan tədqiqatlar və layihələr ölkənin rəqəmsal iqtisadiyyatının inkişafına və müxtəlif sahələrdə innovasiyaların tətbiqinə töhfə verir. 2. Azərbaycanda Süni İntellektin Tarixi İnkişafı Azərbaycanda süni intellektin inkişaf tarixi son 30 ildə formalaşmağa başlamışdır. İlk vaxtlarda bu sahə əsasən akademik mühitdə tədqiqatlarla məhdudlaşdı. 1990-cı illərin əvvəllərində Azərbaycan Milli Elmlər Akademiyasında informasiya texnologiyaları üzrə ilkin araşdırmalar aparılmağa başlandı. 2000-ci illərin əvvəllərində Azərbaycan universitetlərində kompüter elmləri və süni intellekt üzrə təhsil proqramları təqdim edilməyə başlandı. Bakı Dövlət Universiteti və Azərbaycan Texniki Universiteti kimi təhsil müəssisələrində bu sahədə mütəxəssislərin hazırlanmasına başlanılmışdır. 2010-cu illərdə "Azərbaycan 2020: gələcəyə baxış" İnkişaf Konsepsiyasının qəbul edilməsi ilə informasiya texnologiyaları və süni intellekt sahəsində inkişaf daha da sürətləndi.', 'Bu dövrdə həm dövlət, həm də özəl sektorda süni intellekt və maşın öyrənməsi üzrə layihələr həyata keçirilməyə başlandı. 2018-ci ildən etibarən Azərbaycanda "rəqəmsal transformasiya" konsepsiyası çərçivəsində süni intellekt və maşın öyrənməsi texnologiyalarının tətbiqi daha da genişləndi. Bu dövrdə bir sıra startaplar və texnologiya şirkətləri süni intellekt həllərinin yerli bazara təqdim etməyə başladılar. 2020-ci ildən sonra isə Azərbaycanda süni intellekt sahəsində dövlət səviyyəsində strategiyalar hazırlanmağa başlanmışdır. Bu strategiyalar çərçivəsində süni intellekt və maşın öyrənməsi texnologiyalarının müxtəlif

sahələrdə, xüsusilə də dövlət xidmətlərində tətbiqi prioritet olaraq müəyyən edilmişdir.3. Məlumat Elminin Əsasları və Azərbaycanda Tətbiqi Məlumat elmi (Data Science), böyük həcmli məlumatların analizi və onlardan faydalı informasiyanın çıxarılması ilə məşğul olan elm sahəsidir. Bu sahə statistika, riyaziyyat, kompüter elmləri və müxtəlif tətbiq sahələrinin biliklərini birləşdirir.'

- **BGE-M3**

Azərbaycan dili üçün TDE-nin əsas çətinlikləri aşağıdakılardır: 1. Morfoloji mürəkkəblik: Azərbaycan dilində bir sözə çoxlu sayda şəkilçi əlavə edilə bilər, bu da sözün formasını dəyişir. 2. Resursların məhdudluğu: İngilis dili kimi geniş yayılmış dillərlə müqayisədə, Azərbaycan dili üçün annotasiya edilmiş məlumat bazaları və leksik resurslar daha azdır. 3. Standartlaşdırılmış alətlərin azlığı: Azərbaycan dili üçün hazır TDE alətləri və kitabxanaları məhduddur. Bununla belə, son illərdə Azərbaycan dili üçün TDE sahəsində bir sıra tədqiqatlar və layihələr həyata keçirilmişdir: • Azərbaycan dili üçün morfoloji təhlil və lemmalaşdırma alətləri yaradılmışdır. • Azərbaycan dilində mətnlərin avtomatik təsnifatı və sentiment analizi üzrə tədqiqatlar aparılmışdır. • Azərbaycan dilindən digər dillərə və əksinə məşin tərcüməsi sistemləri inkişaf etdirilmişdir. • Azərbaycan dili üçün söz vektorları və dil modelləri yaradılmışdır. 5.3.', 'Təbii Dilin Emalı (Natural Language Processing) və Azərbaycan Dili Təbii dilin emalı (TDE), kompüterlərin insan dilini başa düşmək, emal etmək və generasiya etmək qabiliyyətini təmin edən süni intellekt sahəsidir. Bu sahə, mətnlərin təhlili, məşin tərcüməsi, sentiment analizi, sual-cavab sistemləri və s. kimi tətbiqləri əhatə edir. 5.1. Təbii Dilin Emalının Əsas Komponentləri TDE bir neçə əsas komponenti əhatə edir: • Leksik Analiz: Mətnin sözlərə və cümlələrə bölünməsi. • Morfoloji Analiz: Sözlərin kök və şəkilçilərə ayrılması. • Sintaktik Analiz: Cümlələrin qrammatik strukturunun təhlili. • Semantik Analiz: Mətnin mənasının anlaşılması. • Pragmatik Analiz: Mətnin kontekstdə mənasının başa düşülməsi. 5.2. Azərbaycan Dili üçün Təbii Dilin Emalı Azərbaycan dili, digər türk dilləri kimi, aqlütinativ bir dildir, yəni sözlər kök və şəkilçilərin birləşməsi ilə formalaşır. Bu xüsusiyyət, Azərbaycan dili üçün TDE sistemlərinin yaradılmasında xüsusi yanaşmalar tələb edir.', 'Azərbaycan Dili üçün TDE Layihələri və Tətbiqləri Azərbaycanda TDE sahəsində bir sıra layihələr həyata keçirilmişdir: • Dilmanç: Azərbaycan dili və digər dillər arasında tərcümə xidməti təqdim edən sistem. • AzerBERT: Azərbaycan dili üçün hazırlanmış BERT dil modeli. • AzCorpus: Azərbaycan dilində mətnlərin böyük korpusu. • AzSpell: Azərbaycan dili üçün orfoqrafiya yoxlama sistemi. Bu layihələr, Azərbaycan dilində təbii dilin emalı sahəsində əhəmiyyətli addımlar atmağa imkan vermişdir. Bundan əlavə, Google Translate, Yandex Translate kimi qlobal tərcümə sistemləri də Azərbaycan dilini dəstəkləyir. TDE-nin Azərbaycanda potensial tətbiq sahələri aşağıdakılardır: • E-hökumət xidmətləri: Vətəndaşların sorğularını avtomatik emal etmək və cavablandırmaq üçün. • Təhsil: Azərbaycan dilini öyrənmək və tədris etmək üçün interaktiv alətlər. • Media monitoringi: Sosial mediada və

xəbər saytlarında məzmunun avtomatik təhlili. • Müştəri xidməti: Müştəri sorğularını avtomatik cavablandıran chatbotlar.

- **LaBSE**

Təbii Dilin Emalı (Natural Language Processing) və Azərbaycan Dili Təbii dilin emalı (TDE), kompüterlərin insan dilini başa düşmək, emal etmək və generasiya etmək qabiliyyətini təmin edən süni intellekt sahəsidir. Bu sahə, mətnlərin təhlili, maşın tərcüməsi, sentiment analizi, sual-cavab sistemləri və s. kimi tətbiqləri əhatə edir. 5.1. Təbii Dilin Emalının Əsas Komponentləri TDE bir neçə əsas komponenti əhatə edir: • Leksik Analiz: Mətnin sözlərə və cümlələrə bölünməsi. • Morfoloji Analiz: Sözlərin kök və şəkilçilərə ayrılması. • Sintaktik Analiz: Cümlələrin qrammatik strukturunun təhlili. • Semantik Analiz: Mətnin mənasının anlaşılması. • Pragmatik Analiz: Mətnin kontekstdə mənasının başa düşülməsi. 5.2. Azərbaycan Dili üçün Təbii Dilin Emalı Azərbaycan dili, digər türk dilləri kimi, aqlütinativ bir dildir, yəni sözlər kök və şəkilçilərin birləşməsi ilə formalaşır. Bu xüsusiyyət, Azərbaycan dili üçün TDE sistemlərinin yaradılmasında xüsusi yanaşmalar tələb edir.', 'Azərbaycan dili üçün TDE-nin əsas çətinlikləri aşağıdakılardır: 1. Morfoloji mürekkəblilik: Azərbaycan dilində bir sözə çoxlu sayda şəkilçi əlavə edilə bilər, bu da sözün formasını dəyişir. 2. Resursların məhdudluğu: İngilis dili kimi geniş yayılmış dillərlə müqayisədə, Azərbaycan dili üçün annotasiya edilmiş məlumat bazaları və leksik resurslar daha azdır. 3. Standartlaşdırılmış alətlərin azlığı: Azərbaycan dili üçün hazır TDE alətləri və kitabxanaları məhduddur. Bununla belə, son illərdə Azərbaycan dili üçün TDE sahəsində bir sıra tədqiqatlar və layihələr həyata keçirilmişdir: • Azərbaycan dili üçün morfoloji təhlil və lemmalaşdırma alətləri yaradılmışdır. • Azərbaycan dilində mətnlərin avtomatik təsnifatı və sentiment analizi üzrə tədqiqatlar aparılmışdır. • Azərbaycan dilindən digər dillərə və əksinə maşın tərcüməsi sistemləri inkişaf etdirilmişdir. • Azərbaycan dili üçün söz vektorları və dil modelləri yaradılmışdır. 5.3.', 'Azərbaycan Dili üçün TDE Layihələri və Tətbiqləri Azərbaycanda TDE sahəsində bir sıra layihələr həyata keçirilmişdir: • Dil Mancı: Azərbaycan dili və digər dillər arasında tərcümə xidməti təqdim edən sistem. • AzerBERT: Azərbaycan dili üçün hazırlanmış BERT dil modeli. • AzCorpus: Azərbaycan dilində mətnlərin böyük korpusu. • AzSpell: Azərbaycan dili üçün orfoqrafiya yoxlama sistemi. Bu layihələr, Azərbaycan dilində təbii dilin emalı sahəsində əhəmiyyətli addımlar atmağa imkan vermişdir. Bundan əlavə, Google Translate, Yandex Translate kimi global tərcümə sistemləri də Azərbaycan dilini dəstəkləyir. TDE-nin Azərbaycanda potensial tətbiq sahələri aşağıdakılardır: • E-hökumət xidmətləri: Vətəndaşların sorğularını avtomatik emal etmək və cavablandırmaq üçün. • Təhsil: Azərbaycan dilini öyrənmək və tədris etmək üçün interaktiv alətlər. • Media monitorinqi: Sosial mediada və xəbər saytlarında məzmunun avtomatik təhlili. • Müştəri xidməti: Müştəri sorğularını avtomatik cavablandıran chatbotlar.

Nəticələrin təhlili

3 fərqli modelin 3 fərqli sorğu üzərində qaytardığı nəticələrə əsasən onu deyə bilərik ki, XLM-RoBERTa modeli ən zəif nəticə göstərən modeldir. BGE-M3 və LaBSE modellərinin performansları bir-birinə yaxın olsa da, nisbətən BGE-M3 modelinin nəticələri daha yaxşıdır. İlk 2 sorğu sadə faktiki suallar, sonuncu sorğu isə daha mürəkkəb, cavabı sample dokumentin müxtəlif yerlərində olan sual idi. Buna görə də 3-cü sorğuda ideal cavab olmadı, ancaq LaBSE və BGE-M3 modelləri nisbətən yaxın cavab verməyi bacardı. BGE-M3 modeli ilk 2 sorğuda dəqiq cavab verərkən, LaBSE modeli ilk sorğuya yanlış, 2-ci sorğuya isə düzgün cavab verdi.