

ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆ

ՆԵՐԱԾՈՒԹՅՈՒՆ	4
ԽՆԴԻ ԴՐՎԱԾՔ	5
ԳԼՈՒԽ 1. PDF ՓԱՍՏԱԹՂԹԵՐԻ ԿԱՌՈՒՅՎԱԾՔԸ և ԱՌԱՆՁՆԱՀԱՏԿՈՒԹՅՈՒՆՆԵՐԸ.....	6
1.1. PDF ՖԱՅԼԵՐԻ ԿԱՌՈՒՅՎԱԾՔ	6
1.2. PDF-ի ՇԵՐՏԵՐ	9
1.3. PDF ՓԱՍՏԱԹՂԹԵՐԻՑ ՏԵՔՍՏԻ ԴՈՒՐՍԲԵՐՄԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ	12
1.4. ՏԵՔՍՏԱՅԻՆ ՇԵՐՏԻ ԴՈՒՐՍ ԲԵՐՄԱՆ ԳՐԱԴԱՐԱՆՆԵՐ PYTHON-ՈՒՄ	15
ԳԼՈՒԽ 2. ԱՐՀԵՍՏԱԿԱՆ ԲԱՆԱԿԱՆՈՒԹՅԱՆ ԿԻՐԱՌՈՒԹՅՈՒՆԸ ՏԵՔՍՏԻ ԱՍՓՈՓՄԱՆ ՀԱՄԱՐ.....	19
2.1. ՏԵՔՍՏԻ ԱՍՓՈՓՄԱՆ ՀԻՄՆԱԿԱՆ ՍՈՏԵՑՈՒՄՆԵՐ	19
2.2. ԺԱՄԱՆԱԿԱԿԻՑ ԱՐ ՍՈՂԵԼՆԵՐ ՏԵՔՍՏԻ ԱՍՓՈՓՄԱՆ ՀԱՄԱՐ	24
2.3. ՏԵՔՍՏԻ ԱՍՓՈՓՄԱՆ ԳՆԱՀԱՏՈՒՄ ՄԱՐԴՈՒ ԿՈՂՄԻՑ	27
2.4. ԱՍՓՈՓՄԱՆ ՄԱՐՏԱՀՐԱՎԵՐՆԵՐԸ ԵՐԿԱՐ ՓԱՍՏԱԹՂԹԵՐԻ ՀԱՄԱՐ	29
ԳԼՈՒԽ 3. ՀԱՄԱԿԱՐԳԻ ՆԱԽԱԳԾՈՒՄ ԵՎ ԻՐԱԿԱՆԱՑՈՒՄ	33
3.1. ՀԱՄԱԿԱՐԳԻ ԸՆԴՀԱՆՈՒՐ ՃԱՐՏԱՐԱՊԵՏՈՒԹՅՈՒՆ	33
3.2. PDF ՓԱՍՏԱԹՂԹԵՐԻ ՄՇԱԿՄԱՆ ՍՈՂՈՒԼ	34
3.3. ՏԵՔՍՏԻ ԱՍՓՈՓՄԱՆ ՍՈՂՈՒԼ	36
3.4. ՕԳՏԱԳՈՐԾՈՂԻ ԻՆՏԵՐՖԵՅՍ	38
3.5. API ԻՆՏԵՐՖԵՅՍ	46
3.6. ՕԳՏԱԳՈՐԾՄԱՆ ՕՐԻՆԱԿՆԵՐ	50
ԳԼՈՒԽ 4. ԲՆԱՊԱՀՊԱՆՈՒԹՅՈՒՆ	57
4.1. ԷԼԵԿՏՐՈՆԱԳՆԻՄԱԿԱՆ ԴԱՇՏԻ ԱԶԴԵՑՈՒԹՅՈՒՆԸ ՇՐՋԱԿԱ ՄԻՋԱՎԱՅՐԻ ԵՎ ՄԱՐԴՈՒ ԱՌՈՂՋՈՒԹՅԱՆ ՎՐԱ	57
ԳԼՈՒԽ 5. ԿԵՆՄԱԳՈՐԾՈՒՆԵՈՒԹՅԱՆ ԱՆՎՏԱՆԳՈՒԹՅՈՒՆ	63
5.1. ԷԼԵԿՏՐՈՆԱՅԻՆ ՓԱՍՏԱԹՂԹԵՐԻ ԶԳԱՅՈՒՆ ՏԵՂԵԿԱՏՎՈՒԹՅԱՆ ԱՆՀԱՄԱՆԵԼԻՈՒԹՅՈՒՆ	63
ԳԼՈՒԽ 6. ՏՆՏԵՍԱԳԻՏՈՒԹՅՈՒՆ	68
6.1. ԷԼԵԿՏՐՈՆԱՅԻՆ ՓԱՍՏԱԹՂԹԵՐԻ (PDF) ԲՈՎԱՆԴԱԿՈՒԹՅԱՆ ՃԱՆԱՉՄԱՆ ԵՎ ԱՍՓՈՓՄԱՆ ԱՎՏՈՍՏԱՏՎԱԾ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՄԱՆ ԻՆՔՆԱՐԺԵՔԸ	68
ԵԶՐԱԿԱՑՈՒԹՅՈՒՆՆԵՐ	77
ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ	80

ՆԵՐԱԾՈՒԹՅՈՒՆ

Արդի թվային դարում տեղեկատվության կառավարումը և վերլուծությունը դարձել են կենսական անհրաժեշտություն: Միլիարդավոր փաստաթղթեր են պահպանվում PDF ձևաչափով՝ ընդգրկելով բիզնեսի, գիտության, կրթության և այլ ոլորտների արժեքավոր տեղեկատվություն: Սակայն նման փաստաթղթերի բովանդակության վերլուծությունը հաճախ բարդ է և պահանջում է զգալի ժամանակ և ռեսուրսներ:

Արհեստական բանականությունը (ԱԲ) հնարավորություն է տալիս ստեղծել նորարարական գործիքներ, որոնք լուծում են այս մարտահրավերները՝ ավտոմատացնելով գործընթացները: ԱԲ-ի շնորհիվ հնարավոր է իրականացնել PDF փաստաթղթերի արագ և ճշգրիտ վերլուծություն, առանձնացնել բովանդակության առանցքային հատվածները և ստեղծել հակիրճ ամփոփումներ՝ առանց բովանդակության հիմնական իմաստի կորուստի: Այս մոտեցումը կարևորվում է հատկապես այն դեպքերում, երբ անհրաժեշտ է մշակել մեծ ծավալի տվյալներ, ինչպիսիք են բիզնես հաշվետվությունները, իրավաբանական փաստաթղթերը կամ գիտական հոդվածները:

Բացի դրանից, ԱԲ-ն առաջարկում է նոր հնարավորություններ փաստաթղթերի կառուցվածքային բարդությունների հաղթահարման համար: PDF փաստաթղթերը հաճախ պարունակում են բազմազան տարրեր՝ տեքստ, պատկերներ, աղյուսակներ, գծապատկերներ և գրաֆիկական տարրեր, ինչը դրանց մշակումը դարձնում է բարդ: Այս բոլոր տարրերի միասնական վերլուծությունը պահանջում է ճկուն և խելացի մեթոդաբանություն:

Այս աշխատանքը նպատակ ունի ստեղծել ավտոմատացված համակարգ, որը կկարողանա վերլուծել PDF փաստաթղթերի բովանդակությունը, առանձնացնել կարևոր տեղեկատվությունը և ներկայացնել այն հակիրճ և հասկանալի ձևով: Սա ոչ միայն կբարելավի տվյալների մշակման արդյունավետությունը, այլև կնպաստի ժամանակի խնայմանը և կդարձնի բարդ փաստաթղթերի հետ աշխատելը ավելի դյուրին:

ԽՆԴՐԻ ԴՐՎԱԾՔ

Ներկայումս PDF ձևաչափի փաստաթղթերի վերլուծությունը մեծ մարտահրավեր է, քանի որ դրանք պարունակում են ինչպես տեքստային, այնպես էլ գրաֆիկական տվյալներ, որոնք ավտոմատ մշակման համար բարդություններ են առաջացնում: Խնդիրը ներառում է հետևյալ հիմնական հարցերը.

1. PDF փաստաթղթերի բովանդակության ավտոմատ վերլուծություն, ներառյալ տեքստի և պատկերների ճանաչումը:
2. Կարևոր տեղեկատվության առանձնացում՝ առանց բովանդակության հիմնական իմաստի կորուստի:
3. Ամփոփ տեղեկատվության գեներացիա՝ ապահովելով ստացված արդյունքների հասանելիությունն ու ընթեռնելիությունը:
4. Համակարգի ինտեգրում՝ ապահովելով տարբեր բնագավառներում կիրառելիությունը:

Այս աշխատանքն ուղղված է ԱԲ տեխնոլոգիաների կիրառմանը՝ ստեղծելու ավտոմատացված համակարգ, որը հնարավորություն կտա արդյունավետ մշակել և ամփոփել PDF փաստաթղթերի բովանդակությունը՝ նվազեցնելով մարդկային ներգրավվածությունը և խնայելով ժամանակ ու ռեսուրսներ:

ԳԼՈՒԽ 1. PDF ՓԱՍՏԱԹՂԹԵՐԻ ԿԱՌՈՒՑՎԱԾՔԸ և ԱՌԱՆՁՆԱՀԱՏԿՈՒԹՅՈՒՆՆԵՐԸ

1.1. PDF ՖԱՅԼԵՐԻ ԿԱՌՈՒՑՎԱԾՔ

Portable Document Format (PDF) ձևաչափը մշակվել է Adobe-ի համահիմնադիր Ջոն Ուորնոթի կողմից 1991 թվականին՝ որպես «The Camelot Project»-ի մաս: Նախագծի նպատակն էր ստեղծել ֆայլային ձևաչափ, որը կապահովեր փաստաթղթերի ճշգրիտ ներկայացումը տարբեր համակարգերում՝ անկախ ծրագրային ապահովումից, որով ստեղծվել է փաստաթուղթը: 1993 թվականին Adobe-ը թողարկեց Acrobat 1.0 ծրագիրը և PDF 1.0 ձևաչափը, որը դարձավ PDF-ի առաջին հանրային տարբերակը:

Տարիների ընթացքում PDF ձևաչափը զարգացել է՝ ավելացնելով հնարավորություններ և բարելավելով գոյություն ունեցածները: PDF-ի զարգացման հիմնական փուլերը.

- **PDF 1.0 (1993)** - Հիմնական ֆունկցիոնալություն, ներառյալ տեքստի և պատկերների ներկայացումը
- **PDF 1.1 (1994)** - Ավելացվեցին հղումներ, գույների կառավարում և այլ հնարավորություններ
- **PDF 1.2 (1996)** - Ներդրվեցին ինտերակտիվ ձևեր և JavaScript աջակցություն
- **PDF 1.3 (1999)** - Ավելացվեցին թափանցիկություն և թվային ստորագրություններ
- **PDF 1.4 (2001)** - Ներդրվեցին թափանցիկության խմբեր և XML ձևեր
- **PDF 1.5 (2003)** - Ավելացվեցին շերտեր և բարելավվեց սեղմման տեխնոլոգիան
- **PDF 1.6 (2004)** - Ներդրվեցին AES գաղտնագրում և OpenType տառատեսակներ
- **PDF 1.7 (2006)** - Դարձավ ISO ստանդարտ (ISO 32000-1:2008)
- **PDF 2.0 (2017)** - Թողարկվեց որպես ISO 32000-2:2017, ավելացնելով նոր հնարավորություններ և բարելավելով անվտանգությունը [23]

2008 թվականին PDF 1.7 ձևաչափը ստանդարտացվեց Միջազգային ստանդարտացման կազմակերպության (ISO) կողմից որպես ISO 32000-1:2008: [1] Այսինքն՝ այն այլևս չէր վերահսկվում միայն Adobe-ի կողմից: Այս քայլը նպաստեց PDF-ի ավելի լայն ընդունմանը և զարգացմանը:

PDF ֆայլերի հիմնական բաղադրիչները.

PDF փաստաթուղթը բաղկացած է մի շարք բաղադրիչներից, որոնք ներառում են.

1. **Օբյեկտներ (Objects)** - PDF-ի հիմնական կառուցվածքային միավորները, որոնք կարող են լինել բուլյան արժեքներ, թվեր, տողեր, զանգվածներ, բառարաններ և այլն:
2. **Ֆայլի կառուցվածք (File Structure)** - PDF ֆայլը բաղկացած է չորս հիմնական մասերից.
 - **Գլխագիր (Header)** - Պարունակում է PDF տարբերակի մասին տեղեկատվություն
 - **Մարմին (Body)** - Պարունակում է փաստաթղթի բովանդակությունը ներկայացնող օբյեկտները
 - **Խաչաձև հղումների աղյուսակ (Cross-reference table)** - Պարունակում է հղումներ դեպի ֆայլում գտնվող օբյեկտները
 - **Trailer** - Պարունակում է հղումներ դեպի խաչաձև հղումների աղյուսակը և այլ կարևոր օբյեկտներ
3. **Էջեր և բովանդակություն (Pages and Content)** - PDF փաստաթուղթը կազմված է էջերի հիերարխիկ կառուցվածքով: Յուրաքանչյուր էջ պարունակում է բովանդակության հոսք (content stream), որը նկարագրում է էջի վրա տեղադրված տեքստը, գրաֆիկան և այլ տարրերը:
4. **Ռեսուրսներ (Resources)** - Ներառում են տառատեսակներ, պատկերներ և այլ տարրեր, որոնք օգտագործվում են փաստաթղթի ներկայացման համար:
5. **Ինտերակտիվ տարրեր (Interactive Elements)** - Ներառում են հղումներ, ձևեր, JavaScript սկրիպտներ և այլ ինտերակտիվ տարրեր:
6. **Մետատվյալներ (Metadata)** - Պարունակում են տեղեկատվություն փաստաթղթի մասին, ինչպիսիք են հեղինակը, ստեղծման ամսաթիվը, վերնագիրը և այլն:

PDF փաստաթղթերը կարելի է դասակարգել երկու հիմնական տեսակի՝ թվային (native) և սկանավորված (scanned):

Թվային PDF-ներ

Թվային PDF-ները ստեղծվում են ուղղակիորեն թվային աղբյուրներից, ինչպիսիք են տեքստային խմբագրիչները, էլեկտրոնային աղյուսակները: Այս տեսակի PDF-ները պարունակում են տեքստային շերտ, որը թույլ է տալիս.

- Տեքստի որոնում և ընտրություն
- Տեքստի խմբագրում (եթե թույլատրված է)
- Տեքստի դուրսբերում և մշակում

Թվային PDF-ների առավելություններն են.

- Բարձր որակ և հստակություն
- Փոքր ֆայլի չափ
- Տեքստի որոնման և դուրսբերման հնարավորություն
- Հեշտ խմբագրում և մշակում

Սկանավորված PDF-ներ

Սկանավորված PDF-ները ստեղծվում են ֆիզիկական փաստաթղթերի թվայնացման միջոցով՝ օգտագործելով սկաներ կամ թվային տեսախցիկ: Այս տեսակի PDF-ները հիմնականում պատկերներ են, որոնք չունեն տեքստային շերտ, եթե չի կիրառվել OCR տեխնոլոգիա: Սկանավորված PDF-ների հատկություններն են.

- Տեքստը պատկերի մաս է և ուղղակիորեն հասանելի չէ
- Տեքստի որոնումը և դուրսբերումը պահանջում է OCR տեխնոլոգիաների կիրառում
- Սովորաբար ավելի մեծ ֆայլի չափ
- Որակը կախված է սկանավորման որակից

Սկանավորված PDF-ների մշակման համար անհրաժեշտ է կիրառել OCR տեխնոլոգիաներ, որոնք թույլ են տալիս ճանաչել պատկերում առկա տեքստը և ստեղծել որոնելի տեքստային շերտ: Այս գործընթացը կարող է լինել բարդ և պահանջել լրացուցիչ մշակում՝ կախված սկանավորման որակից, տեքստի լեզվից և այլ գործոններից:

PDF ֆայլերի կառուցվածքի և տեսակների հասկացողությունը կարևոր է PDF փաստաթղթերի բովանդակության ավտոմատ վերլուծության և ամփոփման համակարգի մշակման համար: Տարբեր տեսակի PDF-ների հետ աշխատելու համար անհրաժեշտ է կիրառել տարբեր մոտեցումներ և տեխնոլոգիաներ:

1.2. PDF-ի ՇԵՐՏԵՐ

PDF փաստաթղթերը պարունակում են շերտեր, որոնք ապահովում են բազմաֆունկցիոնալությունը և ճկունությունը: Այս շերտերը թույլ են տալիս պահպանել տեղեկատվությունը և ապահովել ճշգրիտ ներկայացումը: Հիմնական շերտերը ներառում են մետատվյալները, ներդրված օբյեկտները, տեքստային և գրաֆիկական շերտը:

Տեքստային շերտ

Տեքստային շերտը պարունակում է PDF փաստաթղթի տեքստային բովանդակությունը: Թվային PDF-ներում տեքստային շերտը ստեղծվում է ուղղակիորեն փաստաթղթի ստեղծման ժամանակ և պարունակում է տեքստի սիմվոլները, դրանց դիրքը, տառատեսակը և այլ հատկություններ:

Տեքստային շերտի հիմնական բաղադրիչներն են.

1. **Սիմվոլներ (Characters)** - Առանձին տառեր, թվեր և հատուկ նշաններ, որոնք կազմում են տեքստը:
2. **Տառատեսակներ (Fonts)** - Տառատեսակների նկարագրություններ՝ տեքստի ցուցադրման համար: PDF-ը կարող է պարունակել ներդրված տառատեսակներ կամ հղումներ համակարգային տառատեսակներին:
3. **Տեքստի դիրքավորում (Text Positioning)** - Տեղեկատվություն տեքստի դիրքի և չափերի մասին էջի վրա:
4. **Տեքստի հատկություններ (Text Properties)** - Տեքստի գույնը, չափը, ոճը և այլ հատկություններ:

Տեքստային շերտը թույլ է տալիս իրականացնել հետևյալ գործողությունները.

- Տեքստի որոնում և ընտրություն
- Տեքստի պատճենում և դուրսբերում
- Տեքստի խմբագրում (եթե թույլատրված է)
- Տեքստի ինդեքսավորում և որոնում

Սկանավորված PDF-ներում տեքստային շերտը սովորաբար բացակայում է, սակայն, OCR տեխնոլոգիաների կիրառմամբ հնարավոր է ստեղծել տեքստային շերտ, որը կպարունակի ճանաչված տեքստը:

Գրաֆիկական շերտ

Գրաֆիկական շերտը պարունակում է PDF փաստաթղթի գրաֆիկական բովանդակությունը, ներառյալ պատկերները, գծագրերը, գրաֆիկները և այլ վիզուալ տարրերը: Գրաֆիկական շերտը կարող է պարունակել տարբեր տեսակի գրաֆիկական տարրեր.

1. **Ռաստերային պատկերներ (Raster Images)** - Պիքսելային պատկերներ, ինչպիսիք են JPEG, PNG կամ TIFF ձևաչափերը:
2. **Վեկտորային գրաֆիկա (Vector Graphics)** - Մաթեմատիկական բանաձևերով գրաֆիկական տարրեր, որոնք մասշտաբավորվում են առանց որակի կորստի:
3. **Գծեր և ձևեր (Lines and Shapes)** - Պարզ գրաֆիկական տարրեր, ինչպիսիք են գծերը, քառակուսիները, շրջանները և այլն:
4. **Գրադիենտներ և ստվերներ (Gradients and Shadows)** - Գունային անցումներ և ստվերային էֆեկտներ:

Գրաֆիկական շերտի հետ աշխատելիս կարևոր է հաշվի առնել.

- Կարող են պարունակել տեքստ, որը չի հանդիսանում տեքստային շերտի մաս
- Ռաստերային պատկերները ունենան տարբեր որակ և սեղմման մակարդակ
- Վեկտորային գրաֆիկան կարող է պարունակել բարդ կառուցվածքներ և էֆեկտներ

Մետատվյալներ

Մետատվյալները պարունակում են տեղեկատվություն PDF փաստաթղթի մասին, ինչպիսիք են հեղինակը, ստեղծման ամսաթիվը, վերնագիրը, թեմաները և այլն: Մետատվյալները կարևոր են փաստաթղթերի կատալոգավորման, որոնման և կառավարման համար:

PDF-ում մետատվյալները կարող են պահվել տարբեր ձևաչափերով.

1. **Փաստաթղթի տեղեկատվության բառարան (Document Information Dictionary)** - Պարունակում է հիմնական մետատվյալները, ինչպիսիք են հեղինակը, վերնագիրը, ստեղծման ամսաթիվը և այլն:
2. **XMP (Extensible Metadata Platform)** - XML-ի վրա հիմնված ձևաչափ, պահում է ավելի հարուստ և կառուցվածքային մետատվյալներ: [4]

3. **Հատուկ մետատվյալներ (Custom Metadata)** - Օգտագործողի կողմից սահմանված մետատվյալներ, որոնք կարող են պահվել տարբեր ձևաչափերով:
- Մետատվյալները կարող են օգտագործվել հետևյալ նպատակներով.
- Փաստաթղթերի դասակարգում և կատալոգավորում
 - Փաստաթղթերի որոնում և ֆիլտրում
 - Փաստաթղթերի հեղինակության և ամբողջականության ստուգում
 - Փաստաթղթերի պատմության և փոփոխությունների հետևում

Ներդրված օբյեկտներ

PDF փաստաթղթերը կարող են պարունակել ներդրված օբյեկտներ: Այս օբյեկտները կարող են լինել.

1. **Ֆայլեր (Files)** - Ֆայլեր, որոնք կարող են ներդրվել PDF-ում, ինչպիսիք են փաստաթղթերը, էլեկտրոնային աղյուսակները, պրեզենտացիաները և այլն:
 2. **Մուլտիմեդիա (Multimedia)** – Ձայնային ֆայլեր և տեսանյութեր, որոնք կարող են նվազարկվել PDF դիտարկիչի միջոցով:
 3. **3D մոդելներ (3D Models)** - Եռաչափ մոդելներ, որոնք կարող են դիտվել և մանիպուլյացվել PDF դիտարկիչի միջոցով:
 4. **Ինտերակտիվ ձևեր (Interactive Forms)** - Ձևեր, որոնք թույլ են տալիս օգտագործողներին մուտքագրել տվյալներ և փոխազդել փաստաթղթի հետ:
 5. **JavaScript ծրագրեր (JavaScript Scripts)** - Ծրագրեր, որոնք կարող են իրականացնել գործողություններ, ինչպիսիք են տվյալների ստուգումը, հաշվարկները և այլն:
- Ներդրված օբյեկտների հետ աշխատելիս կարևոր առանձնահատկություններ են.
- Կարող են մեծացնել PDF ֆայլի չափը
 - Կարող են պահանջել ծրագրային ապահովում օգտագործման համար
 - Կարող են պարունակել անվտանգության ռիսկեր

PDF-ի շերտերի հասկացողությունը կարևոր է PDF փաստաթղթերի բովանդակության ավտոմատ վերլուծության և ամփոփման համակարգի մշակման համար: Տարբեր շերտերից տեղեկատվության դուրսբերումը և մշակումը պահանջում է տարբեր մոտեցումներ և տեխնոլոգիաներ:

1.3. PDF ՓԱՍՏԱԹՂԹԵՐԻՑ ՏԵՔՍՏԻ ԴՈՒՐՍԲԵՐՄԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Կախված PDF-ի տեսակից և կառուցվածքից, տեքստի դուրսբերման համար կարող են կիրառվել տարբեր տեխնոլոգիաներ:

Տեքստային շերտի ուղղակի դուրսբերում

Տեքստային շերտի ուղղակի դուրսբերումը ներառում է հետևյալ քայլերը.

1. **PDF ֆայլի բացում և վերլուծություն** - PDF ֆայլը բացվում է և վերլուծվում է դրա կառուցվածքը:
2. **Էջերի և բովանդակության հոսքերի բացահայտում** - Բացահայտվում են PDF-ի էջերը և դրանց բովանդակության հոսքերը:
3. **Տեքստային օբյեկտների դուրսբերում** - Բովանդակության հոսքերից դուրս են բերվում տեքստային օբյեկտները:
4. **Տեքստի վերականգնում** - Տեքստային օբյեկտները վերածվում են ընթեռնելի տեքստի՝ հաշվի առնելով դրանց դիրքը, տառատեսակը և այլ հատկություններ: [2]
Տեքստային շերտի ուղղակի դուրսբերման առավելություններն են.
 - Բարձր արագություն և արդյունավետություն
 - Տեքստի ճշգրիտ դուրսբերում
 - Տեքստի հատկությունների պահպանման հնարավորություն
 - Ռեսուրսների օպտիմալ օգտագործումՍակայն, այս մոտեցումն ունի նաև որոշակի սահմանափակումներ.
 - Կիրառելի է միայն թվային PDF-ների համար
 - Կարող է դժվարություններ առաջանալ բարդ կառուցվածքի դեպքում
 - Չի կարող դուրս բերել տեքստը պատկերներից կամ գրաֆիկական տարրերից:

OCR տեխնոլոգիաներ սկանավորված փաստաթղթերի համար

Սկանավորված PDF փաստաթղթերից տեքստի դուրսբերման համար անհրաժեշտ է կիրառել օպտիկական նիշերի ճանաչման (OCR) տեխնոլոգիաներ: OCR-ը թույլ է տալիս ճանաչել պատկերում առկա տեքստը և վերածել այն խմբագրելի և որոնելի տեքստի:

OCR գործընթացը ներառում է հետևյալ հիմնական քայլերը.

1. **Պատկերի նախապատրաստում** - Պատկերը նախապատրաստվում է OCR-ի համար՝ կիրառելով տարբեր ֆիլտրեր և բարելավումներ:
2. **Էջի սեգմենտավորում** - Պատկերը բաժանվում է տարբեր տարրերի՝ տեքստային բլոկներ, պատկերներ, աղյուսակներ և այլն:
3. **Նիշերի ճանաչում** - Տեքստային բլոկներում ճանաչվում են առանձին նիշերը՝ օգտագործելով տարբեր ալգորիթմներ:
4. **Բառերի և նախադասությունների վերականգնում** - Ճանաչված նիշերը միավորվում են բառերի և նախադասությունների:
5. **Հետմշակում** - Ճանաչված տեքստը ենթարկվում է հետմշակման՝ ուղղելու սխալները և բարելավելու որակը:

Ժամանակակից OCR համակարգերը օգտագործում են խորը ուսուցման (deep learning) տեխնոլոգիաներ, ինչպիսիք են կոնվոլյուցիոն նեյրոնային ցանցերը (CNN) և ռեկուրենտ նեյրոնային ցանցերը (RNN):

OCR տեխնոլոգիաների առավելություններն են.

- Հնարավորություն է տալիս դուրս բերել տեքստը սկանավորված փաստաթղթերից
- Աջակցում է բազմաթիվ լեզուների և տառատեսակների
- Կարող է ճանաչել տեքստը նույնիսկ ցածր որակի պատկերներից
- Ստեղծում է որոնելի և խմբագրելի տեքստային շերտ

Սակայն, OCR տեխնոլոգիաներն ունեն նաև որոշակի սահմանափակումներ.

- Ճանաչման ճշգրտությունը կախված է պատկերի որակից
- Պահանջում է զգալի հաշվողական ռեսուրսներ
- Կարող է դժվարություններ առաջանալ բարդ կառուցվածքի դեպքում
- Որոշ լեզուների և տառատեսակների համար ճշգրտությունը կարող է լինել ցածր

Հիբրիդային մոտեցումներ

Շատ դեպքերում PDF փաստաթղթերը կարող են պարունակել և՛ տեքստային շերտ, և՛ պատկերներ, որոնք պարունակում են տեքստ: Այս դեպքերում արդյունավետ է կիրառել հիբրիդային մոտեցումներ, որոնք համադրում են տեքստային շերտի ուղղակի դուրսբերումը և OCR տեխնոլոգիաները:

Հիբրիդային մոտեցումների հիմնական քայլերն են.

1. **PDF-ի տեսակի և կառուցվածքի վերլուծություն** - Վերլուծվում է PDF-ի տեսակը և կառուցվածքը՝ որոշելու համար, թե որ մասերը պարունակում են տեքստային շերտ, և որոնք պահանջում են OCR:
2. **Տեքստային շերտի դուրսբերում** - Դուրս է բերվում տեքստային շերտը այն մասերից, որոնք պարունակում են այն:
3. **Պատկերների դուրսբերում և OCR** - Դուրս են բերվում պատկերները և կիրառվում է OCR:
4. **Արդյունքների միավորում** - Տեքստային շերտից դուրս բերված տեքստը և OCR-ի միջոցով ճանաչված տեքստը միավորվում են:
5. **Հետմշակում և կառուցվածքի վերականգնում** - Միավորված տեքստը ենթարկվում է հետմշակման՝ վերականգնելու փաստաթղթի կառուցվածքը և ուղղելու հնարավոր սխալները:

Հիբրիդային մոտեցումների առավելություններն են.

- Բարձր ճշգրտություն և ամբողջականություն
- Հնարավորություն է տալիս աշխատել տարբեր տեսակի PDF-ների հետ
- Օպտիմալ ռեսուրսների օգտագործում
- Կառուցվածքի ավելի լավ պահպանում

Սակայն, այս մոտեցումներն ունեն նաև որոշակի մարտահրավերներ.

- Բարդ իրականացում և կարգավորում
- Պահանջում է տարբեր տեխնոլոգիաների համադրում
- Կարող է առաջանալ դժվարություններ արդյունքների միավորման ժամանակ

PDF փաստաթղթերից տեքստի դուրսբերման տեխնոլոգիաների ընտրությունը կախված է կոնկրետ խնդրից, փաստաթղթերի տեսակից և պահանջվող ճշգրտությունից: Արդյունավետ համակարգը պետք է ադապտիվ կերպով ընտրի տեխնոլոգիա՝ հիմնվելով փաստաթղթի առանձնահատկությունների վրա:

1.4. ՏԵՔՍՏԱՅԻՆ ՇԵՐՏԻ ԴՈՒՐՍ ԲԵՐՄԱՆ ԳՐԱԴԱՐԱՆՆԵՐ PYTHON-ՈՒՄ

pdfplumber

pdfplumber-ը հզոր գրադարան է PDF-ներից տեքստային տվյալների դուրս բերման համար՝ պահպանելով դասավորությունը և կառուցվածքը: Առանձնահատուկ է աղյուսակային տվյալների ճշգրիտ դուրս բերմամբ: [5]

Հիմնական առանձնահատկությունները

- Տեքստի դասավորության ճշգրիտ պահպանում
- Աղյուսակների հզոր ճանաչում և դուրս բերում
- Տեքստային տարրերի կոորդինատների ստացում
- Գծերի, ուրվագծերի և նկարների դուրս բերում
- Վիզուալիզացիայի աջակցություն

Թերությունները

- Ավելի դանդաղ է քան PyMuPDF
- Մեծ հիշողության օգտագործում բարդ փաստաթղթերի համար
- Պահանջում է լրացուցիչ կախվածություններ

PyMuPDF (fitz)

PyMuPDF-ը (fitz) հիմնված է MuPDF շարժիչի վրա և առաջարկում է բազմաֆունկցիոնալ լուծումներ: Առանձնանում է բարձր արագությամբ և ընդարձակ հնարավորություններով:

Հիմնական առանձնահատկությունները

- Գերազանց արագություն համեմատած այլ գրադարանների հետ
- Բազմաֆորմատ աջակցություն (PDF, XPS, EPUB, FB2, CBZ)
- Տեքստի, պատկերների և մետատվյալների լիարժեք դուրս բերում
- HTML և XML արտահանում
- PDF փաստաթղթերի փոփոխության հնարավորություն
- Տեքստային որոնման գործիքներ, ներառյալ regex որոնում

Թերությունները

- Տեղադրման բարդություն (պահանջում է C գրադարան)
- Բարդ API
- Հնարավոր հիշողության խնդիրներ մեծ փաստաթղթերի համար

pytesseract և OCR գրադարաններ

pytesseract-ը օգտագործում է Tesseract OCR շարժիչը սկանավորված փաստաթղթերից տեքստ դուրս բերելու համար: [7] Կարևոր է հատկապես թվային տեքստային շերտ չունեցող PDF-ների համար:

Հիմնական առանձնահատկությունները

- Սկանավորված փաստաթղթերից տեքստի ճանաչում
- 100+ լեզուների աջակցություն (ներառյալ հայերենը)
- Նկարների նախնական մշակման հնարավորություն
- Բարձր ճշգրտություն նորագույն OCR տեխնոլոգիաներով [8]

Թերությունները

- Պահանջում է Tesseract OCR տեղադրում
- Դանդաղ մշակման արագություն
- Մեծ հիշողության պահանջներ
- Ճշգրտությունը կախված է նկարի որակից

Գրադարանների համեմատական վերլուծություն

Գրադարանների համեմատությունը ներկայացված է աղ. 1.1-ում, աղ. 1.2-ում, աղ. 1.3-ում, աղ. 1.4-ում, աղ. 1.5-ում և աղ. 1.6-ում :

Կիրառման ոլորտներ

- **pdfplumber** - Օպտիմալ է աղյուսակների դուրս բերման և տեքստի դասավորության պահպանման համար
- **PyMuPDF** - Հարմար է բարդ PDF-ների արագ մշակման համար
- **pytesseract** - Անփոխարինելի է սկանավորված փաստաթղթերի համար

Վերջնական ընտրությունը կախված է կոնկրետ խնդրից, մշակվող PDF-ների տեսակից և անհրաժեշտ ճշգրտությունից:

Աղյուսակ 1.1.

Արագության և արդյունավետության համեմատություն:

Գրադարան	Արագություն	Հիշողության օգտագործում	Մշակման ժամանակ
pdfplumber	Միջին	Միջին	Միջին
PyMuPDF	Շատ արագ	Միջին	Արագ
pytesseract	Դանդաղ	Բարձր	Շատ դանդաղ

Աղյուսակ 1.2.

Տեքստի ճշգրտության և ֆորմատավորման համեմատություն:

Գրադարան	Տեքստի ճանաչում	Դասավորության պահպանում	Աղյուսակների աջակցություն	Ֆորմատավորման պահպանում
pdfplumber	Շատ լավ (թվային)	Լավ	Այո	Միջին
PyMuPDF	Գերազանց (թվային)	Լավ	Մասամբ	Լավ
pytesseract	Լավ (սկանավորված)	Թույլ	Մասամբ	Ոչ

Աղյուսակ 1.3.

PDF տեսակների աջակցության համեմատություն:

Գրադարան	Թվային PDF-ներ	Սկանավորված PDF-ներ	Բարդ ֆորմատավորված PDF-ներ	Պաշտպանված PDF-ներ
pdfplumber	Այո	Ոչ	Այո	Մասամբ
PyMuPDF	Այո	Մասամբ	Այո	Այո
pytesseract	Ոչ	Այո	Մասամբ	Ոչ

Աղյուսակ 1.4.

Տեղադրման բարդության և կախվածությունների համեմատություն:

Գրադարան	Տեղադրման բարդություն	Արտաքին կախվածություններ	Օպերացիոն համակարգերի համատեղելիություն
pdfplumber	Միջին	Միջին	Միջին
PyMuPDF	Միջին	Բավականին	Միջին
pytesseract	Բարդ	Պահանջում է Tesseract OCR	Միջին

Աղյուսակ 1.5.

Գործիքների բազմազանության համեմատություն:

Գրադարան	PDF ստեղծում	Խմբագրում	Միավորում և բաժանում	Նկարների դուրս բերում	Մետատվյալների կառավարում
pdfplumber	Ոչ	Ոչ	Ոչ	Այո	Մասամբ
PyMuPDF	Այո	Այո	Այո	Այո	Այո
pytesseract	Ոչ	Ոչ	Ոչ	Ոչ	Ոչ

Աղյուսակ 1.6.

API-ի որակի և փաստաթղթավորման համեմատություն:

Գրադարան	API-ի պարզություն	Փաստաթղթավորում	Համայնքային աջակցություն	Թարմացումների հաճախականություն
pdfplumber	Միջին	Միջին	Միջին	Միջին
PyMuPDF	Ցածր	Շատ լավ	Ակտիվ	Բարձր
pytesseract	Միջին	Միջին	Ակտիվ	Միջին

ԳԼՈՒԽ 2. ԱՐՀԵՍՏԱԿԱՆ ԲԱՆԱԿԱՆՈՒԹՅԱՆ ԿԻՐԱՌՈՒԹՅՈՒՆԸ

ՏԵՔՍՏԻ ԱՄՓՈՓՄԱՆ ՀԱՄԱՐ

2.1. ՏԵՔՍՏԻ ԱՄՓՈՓՄԱՆ ՀԻՄՆԱԿԱՆ ՄՈՏԵՑՈՒՄՆԵՐ

Տեքստի ամփոփումը (text summarization) բնական լեզվի մշակման (NLP) կարևոր խնդիրներից է, որի նպատակն է երկար տեքստից ստեղծել ավելի կարճ տարբերակ՝ պահպանելով հիմնական իմաստը և կարևոր տեղեկատվությունը: Տեքստի ամփոփման հիմնական մոտեցումներն են էքստրակտիվ, աբստրակտիվ և հիբրիդային:

Էքստրակտիվ ամփոփում

Էքստրակտիվ ամփոփումը (extractive summarization) հիմնված է սկզբնական տեքստից առավել կարևոր նախադասությունների կամ հատվածների ընտրության և դրանց միավորման վրա՝ առանց փոփոխությունների: Այս մոտեցումը չի ստեղծում նոր բովանդակություն, այլ ընտրում է գոյություն ունեցող հատվածները:

Էքստրակտիվ ամփոփման հիմնական քայլերը

1. **Տեքստի նախապատրաստում** - Տեքստը բաժանվում է նախադասությունների, կատարվում է տոկենիզացիա, հեռացվում են ստոպ-բառերը և այլն:
2. **Հատվածների կարևորության գնահատում** - Յուրաքանչյուր նախադասության կամ հատվածի համար հաշվարկվում է կարևորության գնահատական՝ օգտագործելով տարբեր մեթոդներ:
3. **Կարևոր հատվածների ընտրություն** - Ընտրվում են առավել բարձր գնահատական ունեցող նախադասությունները կամ հատվածները:
4. **Ամփոփման կազմում** - Ընտրված հատվածները միավորվում են՝ պահպանելով դրանց սկզբնական հերթականությունը:

Էքստրակտիվ ամփոփման մեթոդներ.

1. **Վիճակագրական մեթոդներ**
 - **TF-IDF (Term Frequency-Inverse Document Frequency)** - Որոշում է բառերի կարևորությունը՝ հիմնվելով հաճախականության և եզակիության վրա:

- **TextRank/LexRank** - Գրաֆի վրա հիմնված ալգորիթմներ, որոնք օգտագործում են նախադասությունների միջև նմանությունը:
- **Latent Semantic Analysis (LSA)** - Կիրառում է մատրիցային դեկոմպոզիցիա՝ բացահայտելու թաքնված իմաստային կառուցվածքը:

2. Մեքենայական ուսուցման մեթոդներ

- **Supervised Learning** - Օգտագործում է պիտակավորված տվյալներ՝ սովորելու համար, թե որ նախադասություններն են կարևոր:
- **Unsupervised Learning** - Բացահայտում է կառուցվածքը և կարևոր հատվածները առանց նախապես նշված պիտակների:
- **Reinforcement Learning** - Օպտիմիզացնում է ամփոփման որակը՝ հիմնվելով պարգևատրման ֆունկցիայի վրա:

3. Խորը ուսուցման մեթոդներ

- **CNN (Convolutional Neural Networks)** - Օգտագործվում են տեքստի հատկանիշների դուրսբերման համար:
- **RNN/LSTM/GRU** - Մոդելավորում են տեքստի հաջորդականությունը և կոնտեքստը:
- **BERT և այլ Transformer-ի վրա հիմնված մոդելներ** - Օգտագործում են ուշադրության մեխանիզմներ՝ հասկանալու նախադասությունների կարևորությունը: [9]

Էքստրակտիվ ամփոփման առավելությունները.

- Պահպանում է սկզբնական տեքստի ճշգրտությունը և փաստերը
- Հեշտ է իրականացնել և պահանջում է ավելի քիչ հաշվողական ռեսուրսներ
- Չի ստեղծում քերականական սխալներ կամ անիմաստ նախադասություններ
- Արդյունավետ է փաստերով հարուստ տեքստերի համար

Էքստրակտիվ ամփոփման թերությունները.

- Կարող է ստեղծել ոչ կապակցված կամ կրկնվող ամփոփումներ
- Չի կարող վերաձևակերպել կամ պարզեցնել բարդ գաղափարները
- Սահմանափակված է սկզբնական տեքստի բառապաշարով և ոճով
- Կարող է ներառել ոչ էական մանրամասներ

Աբստրակտիվ ամփոփում

Աբստրակտիվ ամփոփումը (abstractive summarization) ստեղծում է նոր տեքստ: Այս մոտեցումը նման է մարդու կողմից ամփոփմանը, կարող է վերաձևակերպել, պարզեցնել և ընդհանրացնել տեղեկատվությունը:

Աբստրակտիվ ամփոփման հիմնական քայլերը

1. **Տեքստի հասկացողություն** - Համակարգը վերլուծում է սկզբնական տեքստը՝ հասկանալու դրա իմաստը, կառուցվածքը և հիմնական գաղափարները:
2. **Իմաստային ներկայացում** - Ստեղծվում է տեքստի իմաստային ներկայացում, որը կարող է լինել վերացական կառուցվածք, գրաֆ կամ վեկտորային ներկայացում:
3. **Նոր տեքստի գեներացիա** - Իմաստային ներկայացման հիման վրա գեներացվում է նոր տեքստ, որն արտահայտում է հիմնական գաղափարները:

Աբստրակտիվ ամփոփման մեթոդներ

1. **Sequence-to-Sequence մոդելներ**
 - **Encoder-Decoder ճարտարապետություններ** - Կոդավորում են սկզբնական տեքստը և ապակոդավորում են ամփոփումը:
 - **Attention մեխանիզմներ** - Կենտրոնանում են տեքստի կարևոր մասերի վրա:
 - **Pointer-Generator Networks** - Համադրում են նոր բառերի գեներացիան և սկզբնական տեքստից բառերի պատճենումը:
2. **Transformer-ի վրա հիմնված մոդելներ**
 - **BART (Bidirectional and Auto-Regressive Transformers)** - Նախատեսված է տեքստի գեներացիայի խնդիրների համար: [10]
 - **T5 (Text-to-Text Transfer Transformer)** - Ներկայացնում է խնդիրները տեքստից տեքստ փոխակերպումներով:
 - **GPT (Generative Pre-trained Transformer)** - Օգտագործում է ավտոռեգրեսիվ մոդելավորում տեքստի գեներացիայի համար:
3. **Նախապես ուսուցանված լեզվական մոդելներ**
 - **Fine-tuning** - Նախապես ուսուցանված մոդելների հարմարեցում:
 - **Few-shot learning** - Մոդելների ուսուցանում սահմանափակ օրինակներով:
 - **Zero-shot learning** - Ամփոփումների ստեղծում առանց կոնկրետ ամփոփման տվյալների վրա ուսուցանման:

Աբստրակտիվ ամփոփման առավելությունները.

- Ստեղծում է ավելի բնական և կապակցված ամփոփումներ
- Կարող է վերաձևակերպել և պարզեցնել բարդ գաղափարները
- Ունակ է ընդհանրացնել և միավորել տեղեկատվությունը տարբեր մասերից
- Կարող է ստեղծել ավելի կարճ և տեղեկատվական ամփոփումներ

Աբստրակտիվ ամփոփման թերությունները.

- Կարող է ստեղծել փաստացի սխալներ կամ հորինել տեղեկատվություն
- Պահանջում է զգալի հաշվողական ռեսուրսներ
- Ավելի բարդ է իրականացնել և ուսուցանել
- Կարող է ստեղծել քերականական սխալներ կամ անիմաստ նախադասություններ

Հիբրիդային մոտեցումներ

Հիբրիդային մոտեցումները համադրում են էքստրակտիվ և աբստրակտիվ ամփոփման առավելությունները:

Հիբրիդային ամփոփման հիմնական ռազմավարություններ

1. Երկփուլ մոտեցում

- Առաջին փուլում կիրառվում է էքստրակտիվ ամփոփում՝ ընտրելու կարևոր հատվածները
- Երկրորդ փուլում կիրառվում է աբստրակտիվ ամփոփում՝ ընտրված հատվածների հիման վրա ստեղծելու վերջնական ամփոփումը

2. Ուղղորդված աբստրակտիվ ամփոփում

- Էքստրակտիվ մեթոդները օգտագործվում են կարևոր տեղեկատվության բացահայտման համար
- Այս տեղեկատվությունն օգտագործվում է աբստրակտիվ մոդելի ուղղորդման համար

3. Բազմամակարդակ ամփոփում

- Տարբեր մակարդակներում կիրառվում են տարբեր մոտեցումներ (նախադասություն, պարբերություն, բաժին)
- Արդյունքները միավորվում են՝ ստեղծելու բովանդակ ամփոփում

Հիբրիդային մոտեցումների առավելությունները

- Բարձր ճշգրտություն և ամբողջականություն
- Հավասարակշռություն փաստացի ճշգրտության և բնականության միջև
- Արդյունավետ ռեսուրսների օգտագործում
- Հարմարեցվելիություն տարբեր տեսակի տեքստերի համար

Հիբրիդային մոտեցումների մարտահրավերները

- Բարդ իրականացում և կարգավորում
- Պահանջում է տարբեր մոդելների համադրում
- Կարող է առաջանալ դժվարություններ արդյունքների ինտեգրման ժամանակ

Տեքստի ամփոփման մոտեցման ընտրությունը կախված է կոնկրետ խնդրից, տեքստի տեսակից, պահանջվող ամփոփման որակից և հասանելի ռեսուրսներից: PDF փաստաթղթերի բովանդակության ավտոմատ վերլուծության և ամփոփման համակարգի մշակման ժամանակ կարևոր է հաշվի առնել այս գործոնները և ընտրել համապատասխան մոտեցումը կամ դրանց համադրությունը:

2.2. ԺԱՄԱՆԱԿԱԿԻՑ ԱՐ ՄՈՂԵԼՆԵՐ ՏԵՔՍՏԻ ԱՄՓՈՓՄԱՆ ՀԱՄԱՐ

Վերջին տարիներին արհեստական բանականության և խորը ուսուցման ոլորտում առաջընթացը հանգեցրել է տեքստի ամփոփման համար նախատեսված հզոր մոդելների զարգացմանը:

Transformer-ի հիման վրա հիմնական մոդելներ

Transformer ճարտարապետությունը, որը ներկայացվել է 2017 թվականին "Attention is All You Need" հոդվածում, հեղափոխություն կատարեց բնական լեզվի մշակման ոլորտում: Այն հիմնված է ուշադրության (attention) մեխանիզմների վրա, որոնք թույլ են տալիս մոդելին կենտրոնանալ տեքստի տարբեր մասերի վրա՝ հասկանալու համար դրանց միջև կապերը: Այս ճարտարապետության հիման վրա զարգացել են մի շարք հայտնի մոդելներ՝

- **BERT** (Bidirectional Encoder Representations from Transformers) - Google-ի կողմից մշակված երկկողմանի կոդավորիչ, որը հատկապես արդյունավետ է էքստրակտիվ ամփոփման համար:
- **GPT** (Generative Pre-trained Transformer) - OpenAI-ի կողմից մշակված ավտոռեգրեսիվ մոդել, որն ունի հզոր տեքստի գեներացիայի հնարավորություններ:
- **T5** (Text-to-Text Transfer Transformer) - Google-ի կողմից մշակված մոդել, որը ներկայացնում է բոլոր բնական լեզվի մշակման խնդիրները որպես տեքստից տեքստ փոխակերպումներ:
- **BART** (Bidirectional and Auto-Regressive Transformers) - Facebook AI Research-ի կողմից մշակված մոդել, որը համադրում է BERT-ի երկկողմանի կոդավորիչը և GPT-ի ավտոռեգրեսիվ ապակոդավորիչը:

Gemma 3 մոդելի առանձնահատկությունները

Gemma 3-ը Google DeepMind-ի կողմից մշակված բաց կոդով լեզվական մոդել է, որը հանդիսանում է տեքստի ամփոփման և այլ բնական լեզվի մշակման խնդիրների համար

հզոր գործիք: Այն թողարկվել է 2024 թվականին որպես Gemini մոդելների ընտանիքի բաց կոդով տարբերակ:

Ճարտարապետություն և նախապես ուսուցանում

Gemma 3-ը հիմնված է Transformer ապակոդավորիչի ճարտարապետության վրա, ինչպես GPT մոդելները, սակայն ունի մի շարք առանձնահատկություններ՝

- **Հարմարեցված ուշադրության մեխանիզմներ** - Գեմա 3-ը օգտագործում է բարելավված ուշադրության մեխանիզմներ, որոնք թույլ են տալիս ավելի արդյունավետ մշակել երկար տեքստեր:
- **Նախապես ուսուցանման մեթոդաբանություն** - Մոդելը նախապես ուսուցանվել է ինչպես սովորական տեքստի կանխատեսման, այնպես էլ հատուկ մշակված խնդիրների վրա, որոնք օպտիմալացված են հրահանգների կատարման համար:
- **Համեմատաբար փոքր չափ** - Gemma 3-ը հասանելի է տարբեր չափերի, ինչը թույլ է տալիս այն օգտագործել նույնիսկ սահմանափակ ռեսուրսներով սարքերում:

Gemma 3-ի առավելությունները տեքստի ամփոփման համար

Gemma 3-ը ունի մի շարք առավելություններ, որոնք դարձնում են այն հատկապես արդյունավետ տեքստի ամփոփման խնդիրների համար՝

1. **Հզոր կոնտեքստի հասկացողություն** - Մոդելը կարող է հասկանալ և պահպանել երկար տեքստերի կոնտեքստը, ինչը էական է ամփոփման համար:
2. **Աբստրակտիվ ամփոփման հնարավորություններ** - Gemma 3-ը կարող է ստեղծել բարձրորակ աբստրակտիվ ամփոփումներ, որոնք պահպանում են սկզբնական տեքստի հիմնական գաղափարները՝ միաժամանակ վերաձևակերպելով դրանք:
3. **Բազմալեզու հնարավորություններ** - Չնայած նախապես ուսուցանվել է հիմնականում անգլերենով, Gemma 3-ը ցուցաբերում է լավ արդյունքներ նաև այլ լեզուներով:
4. **Բաց կոդ և հարմարեցման ճկունություն** - Մոդելի բաց կոդով բնույթը թույլ է տալիս հարմարեցնել այն կոնկրետ խնդիրների և ոլորտների համար:
5. **Արդյունավետություն հաշվողական ռեսուրսների տեսանկյունից** - Համեմատած նույնատիպ այլ մոդելների հետ, Gemma 3-ը պահանջում է համեմատաբար քիչ հաշվողական ռեսուրսներ:

Gemma 3-ի կիրառությունը տեքստի ամփոփման համար

Տեքստի ամփոփման համար Gemma 3-ը կարելի է կիրառել մի շարք եղանակներով՝

1. **Zero-shot ամփոփում** - Մոդելին տրամադրվում է պարզապես հրահանգ և սկզբնական տեքստ, առանց որևէ լրացուցիչ օրինակների:
2. **Few-shot ամփոփում** - Մոդելին տրամադրվում են մի քանի օրինակներ «սկզբնական տեքստ-ամփոփում» զույգերի տեսքով, որոնք օգնում են ուղղորդել մոդելի վարքը:
3. **Fine-tuned Gemma 3** - Մոդելը հարմարեցվում է կոնկրետ ամփոփման խնդրի համար՝ օգտագործելով տվյալների հավաքածու, որը պարունակում է «սկզբնական տեքստ-ամփոփում» զույգեր:
4. **Gemma 3 + Ոլորտային հարմարեցում** - Մոդելը նախ հարմարեցվում է կոնկրետ ոլորտի տեքստերի վրա, այնուհետև հատուկ ամփոփման տվյալների վրա:

Մոդելների համեմատական առավելություններ

Տեքստի ամփոփման համար նախատեսված տարբեր մոդելների համեմատությունը ցույց է տալիս, որ՝

- **BERT**-ը լավ ընտրություն է էքստրակտիվ ամփոփման համար, երբ կարևոր է տեքստի ճշգրիտ հասկացողությունը:
- **GPT** մոդելները հզոր են աբստրակտիվ ամփոփման մեջ, սակայն պահանջում են զգալի հաշվողական ռեսուրսներ:
- **T5** և **BART** մոդելները համարվում են հավասարակշռված լուծումներ, որոնք արդյունավետ են տարբեր տիպի ամփոփման խնդիրների համար:
- **Gemma 3**-ը առաջարկում է օպտիմալ հաշվեկշիռ արդյունավետության և հաշվողական ռեսուրսների միջև, ինչպես նաև բաց կոդով բնույթ, որը թույլ է տալիս հեշտությամբ հարմարեցնել այն տարբեր խնդիրների համար:

Այսպիսով, Gemma 3-ը ընտրվել է որպես օպտիմալ մոդել՝ հաշվի առնելով դրա հզոր տեքստի ամփոփման հնարավորությունները, բաց կոդով բնույթը, համեմատաբար ցածր հաշվողական պահանջները և բարձր հարմարեցման ճկունությունը:

2.3. ՏԵՔՍՏԻ ԱՄՓՈՓՄԱՆ ԳՆԱՀԱՏՈՒՄ ՄԱՐԴՈՒ ԿՈՂՄԻՑ

Չնայած ավտոմատ մետրիկաների զարգացմանը, մարդու կողմից իրականացվող գնահատումը շարունակում է մնալ տեքստի ամփոփման որակի գնահատման "ոսկե ստանդարտը": Այն թույլ է տալիս գնահատել այնպիսի ասպեկտներ, որոնք դժվար է չափել ավտոմատ կերպով:

Մարդու կողմից գնահատման մեթոդները

1. **Որակական գնահատում՝** փորձագետները գնահատում են ամփոփումների որակը՝ հիմնվելով այնպիսի չափանիշների վրա, ինչպիսիք են՝
 - Տեղեկատվության ամբողջականությունը
 - Կարևոր տեղեկատվության ներառումը
 - Կրկնության բացակայությունը
 - Ընթերցելիությունը և կապակցվածությունը
 - Քերականական ճշգրտությունը
2. **Քանակական գնահատում՝** փորձագետները տալիս են թվային գնահատականներ կամ դասակարգում են ամփոփումները՝ հիմնվելով տարբեր չափանիշների վրա:
3. **Համեմատական գնահատում՝** գնահատողները համեմատում են տարբեր համակարգերի ամփոփումները մեկը մյուսի կամ ստուգիչ ամփոփումների հետ:

Գնահատման չափանիշներ

1. **Բովանդակային համապատասխանություն՝** որքանով է ամփոփումը ներառում բնօրինակի կարևոր տեղեկատվությունը
2. **Ոճական և քերականական ճշգրտություն՝** ամփոփման լեզվական որակը
3. **Կապակցվածություն և կոհերենտություն՝** որքանով է ամփոփումը հասկանալի և տրամաբանորեն կազմակերպված
4. **Ընդհանուր որակ՝** ամփոփման ընդհանուր օգտակարությունն օգտագործողի համար [12]

Մարդու կողմից գնահատման առավելությունները և թերությունները

Առավելություններ

- Հաշվի է առնում տեքստի բարդ ասպեկտներ, որոնք դժվար է չափել ավտոմատ կերպով
 - Կարող է գնահատել սուբյեկտիվ հատկանիշներ, ինչպիսիք են հասկանալիությունը և կապակցվածությունը
 - Արտացոլում է ամփոփման իրական արժեքը վերջնական օգտագործողի համար
- Թերություններ**
- Ժամանակատար և թանկ է
 - Սուբյեկտիվ է և կարող է ունենալ գնահատողների միջև տարբերություններ
 - Դժվար է իրականացնել մեծ թվով ամփոփումների համար
 - Հնարավոր չէ կիրառել իրական ժամանակում

Գործնականում, տեքստի ամփոփման համակարգերի արդյունավետ գնահատման համար հաճախ կիրառվում են համակցված մոտեցումներ, որոնք միավորում են և՛ ավտոմատ մետրիկաները, և՛ մարդու կողմից իրականացվող գնահատումը: Այս մոտեցումը թույլ է տալիս արագ գտել անարդյունավետ լուծումները ավտոմատ մետրիկաների միջոցով, իսկ այնուհետև կիրառել ավելի խորը մարդկային գնահատում առավել հեռանկարային տարբերակների համար: [11]

Տեքստի ամփոփման գնահատման մեթոդների ուսումնասիրությունը ցույց է տալիս, որ չկա մեկ համընդհանուր մետրիկա, որը կարող է արտացոլել ամփոփման բոլոր կարևոր ասպեկտները: Մարդու կողմից գնահատումը շարունակում է մնալ կարևորագույն մեթոդ, հատկապես տեքստի բարդ հատկությունները գնահատելիս:

2.4. ԱՄՓՈՓՄԱՆ ՄԱՐՏԱՀՐԱՎԵՐՆԵՐԸ ԵՐԿԱՐ ՓԱՍՏԱԹՂԹԵՐԻ ՀԱՄԱՐ

Երկար փաստաթղթերի ամփոփումը հանդիսանում է բարդ և բազմաշերտ խնդիր, որը պահանջում է հատուկ մոտեցումներ և մեթոդներ:

Կոնտեքստի պահպանման խնդիրներ

Երկար փաստաթղթերի ամփոփման ժամանակ կոնտեքստի պահպանումը հանդիսանում է առանցքային խնդիր: Առկա մարտահրավերները.

Իմաստային կապերի պահպանում

Երկար փաստաթղթերում հաճախ առկա են իմաստային բարդ կապեր տարբեր հատվածների միջև: Օրինակ՝ ներածության մեջ նշված գաղափարները կարող են զարգանալ և ամբողջանալ հետագա բաժիններում: Ամփոփման ժամանակ այս կապերի խզումը կարող է հանգեցնել փաստաթղթի ընդհանուր իմաստի աղավաղման: [14]

Համատեքստային հղումների խնդիր

Փաստաթղթերը հաճախ պարունակում են ներքին հղումներ, որոնք կապակցում են տարբեր մասեր: Օրինակ՝ "ինչպես նշվեց 2-րդ բաժնում" կամ "համաձայն վերը նշված վերլուծության": Ամփոփման ժամանակ այս հղումները կարող են կորցնել իրենց համատեքստը և դառնալ անհասկանալի:

Հեղինակային նպատակի պահպանում

Յուրաքանչյուր փաստաթուղթ ունի հեղինակային որոշակի նպատակ և ուղղվածություն: Կոնտեքստի պահպանման մարտահրավերը ներառում է նաև այս ընդհանուր նպատակի և տոնայնության ճիշտ փոխանցումը ամփոփման մեջ:

Մասնագիտական տերմինաբանության համատեքստ

Մասնագիտական փաստաթղթերում տերմինները հաճախ ներկայացվում են որոշակի համատեքստում, որը կարող է ամբողջ փաստաթղթի տարբեր հատվածներում լինել: Ամփոփման ժամանակ այս համատեքստի կորուստը կարող է հանգեցնել տերմինների սխալ մեկնաբանման: [15]

Փաստաթղթի մասնատման ռազմավարություններ

Երկար փաստաթղթերի արդյունավետ ամփոփման համար կարևոր է մշակել մասնատման ռազմավարություններ: Առկա մոտեցումները և մարտահրավերները.

Իմաստային մասնատում

Իմաստային մասնատումը ենթադրում է փաստաթղթի բաժանում ըստ թեմատիկ կամ գաղափարական միավորների: Հաճախ այս միավորները հստակ սահմանազատված չեն և ունեն փոխկապակցվածություններ: [16]

Առավելությունները

- Պահպանվում են իմաստային կապերը
- Ամփոփումը ստանում է ավելի կոհերենտ կառուցվածք

Մարտահրավերները

- Պահանջում է խորը բովանդակային վերլուծություն
- Սուբյեկտիվ մոտեցման ռիսկ

Կառուցվածքային մասնատում

Այս մոտեցումը հիմնվում է փաստաթղթի ֆորմալ կառուցվածքի վրա՝ բաժանելով այն ըստ գլուխների, բաժինների և ենթաբաժինների:

Առավելությունները

- Հեշտ իրականացվող և օբյեկտիվ մեթոդ
- Պահպանում է փաստաթղթի նախնական կառուցվածքը

Մարտահրավերները

- Կարող է խաթարել թեմատիկ կապերը
- Ոչ միշտ է արտացոլում իմաստային կարևորությունը

Քանակական մասնատում

Փաստաթուղթը կարելի է մասնատել պարզապես ըստ քանակական չափանիշների՝ էջերի քանակի, նախադասությունների կամ բառերի թվի:

Առավելությունները

- Հեշտ ավտոմատացվող
- Չի պահանջում բովանդակային վերլուծություն

Մարտահրավերները

- Իմաստային կապերի խաթարման բարձր ռիսկ

- Անտեսում է տեքստի սեմանտիկ կառուցվածքը

Հիբրիդային մոտեցումներ

Գործնականում հաճախ օգտագործվում են վերոնշյալ մոտեցումների համադրումներ: Օրինակ՝ նախնական կառուցվածքային մասնատումը, այնուհետև յուրաքանչյուր հատվածի իմաստային վերլուծությունը:

Հիերարխիկ ամփոփման մոտեցումներ

Հիերարխիկ ամփոփումը հանդիսանում է երկար փաստաթղթերի արդյունավետ մշակման ռազմավարություն, որը ենթադրում է բազմամակարդակ մոտեցում:

"Վերնից ներքև" մոտեցում

Այս մեթոդը սկսվում է ընդհանուր, բարձր մակարդակի ամփոփումից, որը հետագայում մանրամասնվում է ըստ անհրաժեշտության:

Գործընթացը

1. Փաստաթղթի համընդհանուր թեմայի և հիմնական գաղափարների որոշում
2. Առանցքային բաժինների բացահայտում
3. Յուրաքանչյուր բաժնի աստիճանաբար մանրամասնեցում

Մարտահրավերները

- Նախնական փուլում կարող են բաց թողնվել կարևոր մանրամասներ
- Պահանջում է փաստաթղթի ամբողջական ընկալում սկզբից

"Ներքևից վերև" մոտեցում

Այս մեթոդը սկսվում է փաստաթղթի փոքր հատվածների մանրամասն ամփոփումից և աստիճանաբար միավորում դրանք ավելի բարձր մակարդակների:

Գործընթացը

1. Փաստաթղթի մասնատում փոքր միավորների
2. Յուրաքանչյուր միավորի առանձին ամփոփում
3. Միավորների աստիճանական համախմբում՝ ստեղծելով ավելի ընդհանուր ամփոփումներ

Մարտահրավերները

- Կարող է բերել կրկնությունների

- Ռիսկ կա կորցնելու ընդհանուր պատկերը

Հիերարխիկ ամփոփման ավտոմատացումը

Արհեստական բանականության և մեքենայական ուսուցման զարգացումը թույլ է տալիս աստիճանաբար ավտոմատացնել հիերարխիկ ամփոփման գործընթացը:

Ժամանակակից մոտեցումները

- Տրանսֆորմերների վրա հիմնված մոդելներ, որոնք կարող են մշակել երկար տեքստեր
- Ռեկուրսիվ նեյրոնային ցանցեր, հիերարխիկ կառուցվածքների համար
- Հիբրիդային համակարգեր, որոնք համադրում են ավտոմատ և մարդու կողմից վերահսկվող մեթոդները

Սահմանափակումները

- Սեմանտիկ իմաստի պահպանման խնդիրներ
- Մարդու լեզվի բարդ նրբությունների ըմբռնման սահմանափակումներ
- Մասնագիտական տերմինաբանության և կոնտեքստի ճիշտ մեկնաբանման դժվարություններ

Երկար փաստաթղթերի ամփոփումը պահանջում է համակարգված մոտեցում, որը հաշվի է առնում ինչպես կոնտեքստի պահպանման խնդիրները, այնպես էլ մասնատման և հիերարխիկ կազմակերպման ռազմավարությունները: Ժամանակակից տեխնոլոգիաների և մեթոդների զարգացումը թույլ է տալիս աստիճանաբար հաղթահարել այս մարտահրավերները՝ ապահովելով ավելի արդյունավետ և ճշգրիտ ամփոփումներ:

Այնուամենայնիվ, մարդու դերը այս գործընթացում շարունակում է մնալ առանցքային, հատկապես երբ խոսքը վերաբերում է բարդ գիտական, իրավական կամ տեխնիկական փաստաթղթերին, որոնք պահանջում են մասնագիտական գիտելիքներ և բովանդակության խորը ըմբռնում:

ԳԼՈՒԽ 3. ՀԱՄԱԿԱՐԳԻ ՆԱԽԱԳԾՈՒՄ ԵՎ ԻՐԱԿԱՆԱՑՈՒՄ

3.1. ՀԱՄԱԿԱՐԳԻ ԸՆԴՀԱՆՈՒՐ ՃԱՐՏԱՐԱՊԵՏՈՒԹՅՈՒՆ

Համակարգը բաղկացած է հետևյալ հիմնական բաղադրիչներից՝

- PDF մշակման մոդուլ (pdfplumber և PyMuPDF գրադարաններով)
- Պատկերների ճանաչման մոդուլ (OCR՝ pytesseract գրադարանով)
- Տեքստի ամփոփման և դասակարգման մոդուլ (Google Gemma 3 մոդելի օգտագործմամբ)

Համակարգը կառուցված է մոդուլային սկզբունքով՝ առանձին բաղադրիչները կարող են գործել անկախ, բայց փոխկապակցված են ընդհանուր աշխատանքային հոսքում:

Համակարգում տվյալների հոսքը հետևյալն է՝

1. PDF փաստաթղթի ներմուծում համակարգ
2. Տեքստի և պատկերների առանձնացում փաստաթղթից
3. Պատկերների վրա OCR կիրառում
4. Տեքստային և OCR-ից ստացված տվյալների միավորում ըստ էջերի և դիրքերի
5. Ամբողջական տեքստի ամփոփում AI մոդելի օգնությամբ
6. Ամփոփված տեքստի դասակարգում նախասահմանված դասերի համաձայն

Համակարգում օգտագործվել են հետևյալ հիմնական տեխնոլոգիաները և գրադարանները՝

- **pdfplumber** - PDF-ից տեքստային տվյալների առանձնացման համար
- **PyMuPDF (fitz)** - PDF-ից պատկերների և դրանց դիրքերի առանձնացման համար
- **Pillow (PIL)** - պատկերների մշակման համար
- **pytesseract** - պատկերներից տեքստի ճանաչման (OCR) համար
- **Google Genai** - տեքստի ամփոփման և դասակարգման համար AI մոդելների կիրառմամբ
- **Gemma 3** - Google-ի կողմից մշակված լեզվական մոդել, որն օգտագործվում է տեքստի ամփոփման և դասակարգման համար

3.2. PDF ՓԱՍՏԱԹՂԹԵՐԻ ՄՇԱԿՄԱՆ ՄՈՂՈՒԼ

Փաստաթղթերի ներբեռնում և նախնական մշակում

Համակարգը սկսում է PDF փաստաթղթի ներբեռնումով: Համակարգը կարող է աշխատել ինչպես տեքստային, այնպես էլ սկանավորված PDF փաստաթղթերի հետ:

Փաստաթղթի նախնական մշակման գործընթացը ներառում է՝

- Ֆայլի բացում և կարդում pdfplumber և PyMuPDF գրադարանների միջոցով
- Փաստաթղթի էջերի հերթական մշակում
- Յուրաքանչյուր էջի բովանդակության տարանջատում տեքստի և պատկերների

Տեքստի և պատկերների դուրս բերում

Համակարգում տեքստի և պատկերների դուրս բերումն իրականացվում է գուգահեռաբար երկու գրադարանների միջոցով՝

Տեքստի դուրս բերում

- pdfplumber գրադարանի միջոցով համակարգն առանձնացնում է տեքստը յուրաքանչյուր էջից
- Տեքստը պահպանվում է extracted_parts ցուցակում՝ էջի համարի հետ միասին
- Եթե էջը տեքստ չի պարունակում, ապա պահպանվում է դատարկ տող

Պատկերների դուրս բերում

- PyMuPDF (fitz) գրադարանի միջոցով համակարգն առանձնացնում է պատկերները յուրաքանչյուր էջից
- Յուրաքանչյուր պատկերի համար որոշվում է դրա դիրքը էջում (y_pos), ինչը հետագայում օգտագործվում է պատկերներից առանձնացված տեքստը ճիշտ դիրքում տեղադրելու համար
- Պատկերները պահպանվում են առանձին ֆայլերում՝ համապատասխան էջի և հերթական համարի նշումով

OCR կիրառում սկանավորված փաստաթղթերի համար

Համակարգում OCR (Optical Character Recognition) տեխնոլոգիայի կիրառումն իրականացվում է pytesseract գրադարանի միջոցով:

- Յուրաքանչյուր առանձնացված պատկեր ենթարկվում է OCR-ի
- OCR-ի համար օգտագործվում են հայերեն, ռուսերեն և անգլերեն լեզուների լեզվական փաթեթները (`lang='hye+rus+eng'`)
- Պատկերներից ստացված տեքստը պահպանվում է `image_data` ցուցակում՝ էջի համարի և պատկերի դիրքի հետ միասին

Սա հատկապես կարևոր է սկանավորված կամ պատկերներ պարունակող PDF փաստաթղթերի համար, որոնցում տեքստը հնարավոր չէ ուղղակիորեն առանձնացնել:

3.3. ՏԵՔՍՏԻ ԱՄՓՈՓՄԱՆ ՄՈՂՈՒԼ

Տեքստի նախապատրաստում և մաքրում

Համակարգում տեքստի նախապատրաստումը և մաքրումն իրականացվում է հետևյալ քայլերով՝

- PDF-ից առանձնացված տեքստային տվյալների և OCR-ից ստացված տեքստի համադրում ըստ էջերի և պատկերների դիրքերի
- Տեքստային և OCR տվյալների ճիշտ հերթականության ապահովում
- Միավորված տեքստի ստացում՝ պահպանելով բովանդակության իմաստային հաջորդականությունը

Համակարգն իրականացնում է հետևյալ գործընթացը՝

1. Տեքստային տվյալները տեղադրվում են ըստ էջերի հերթականության
2. Յուրաքանչյուր էջում OCR-ից ստացված տեքստը տեղադրվում է այն դիրքում, որտեղ պատկերը գտնվում էր բնօրինակ փաստաթղթում
3. Արդյունքում ստացվում է մեկ միասնական տեքստ (`full_text` փոփոխականում), որը պահպանում է փաստաթղթի բովանդակության ամբողջականությունը

Ամփոփման մոդելի ընտրություն և կարգավորում

Համակարգում ամփոփման համար օգտագործվում է Google-ի Gemma 3 մոդելը՝

- Օգտագործվում է "gemma-3-27b-it" մոդելը, որը Google-ի կողմից մշակված հզոր լեզվական մոդել է
- Այս մոդելը կարող է աշխատել տարբեր լեզուներով, ներառյալ հայերենը
- Ամփոփման հարցումը (prompt) ձևակերպվում է այնպես, որ պահպանվի փաստաթղթի բնօրինակ լեզուն

Համակարգում ամփոփման հարցումն ունի հետևյալ տեսքը՝

“please do summarization for this text {full_text} on same language of document (if text in Armenian do it on Armenian), without comments from you”

Ամփոփման գործընթացի իրականացում

Ամփոփման գործընթացն իրականացվում է հետևյալ քայլերով՝

1. Google Genai API-ին միանալու համար ստեղծվում է հաճախորդ
2. Նախապատրաստված տեքստը (full_text) ուղարկվում է ամփոփման համար
3. Ստացված ամփոփումը պահպանվում է summarization_response փոփոխականում
4. Ամփոփված տեքստն օգտագործվում է հետագա դասակարգման համար

Բացի ամփոփումից, համակարգը նաև իրականացնում է փաստաթղթի դասակարգում նախապես սահմանված կատեգորիաների միջոցով (ֆինանսական, իրավաբանական, վարչական, տեխնիկական, անձնական, ակադեմիական, գովազդային, բժշկական): Դասակարգման համար օգտագործվում է նույն Gemma 3 մոդելը, որն ընտրում է համապատասխան կատեգորիան ըստ ամփոփված տեքստի բովանդակության:

Համակարգում դասակարգման հարցումն ունի հետևյալ տեսքը՝

"please do classification for this text {summarization_response.text} using only this classes {classes}, without comments from you"

Այսպիսով, համակարգն ապահովում է PDF փաստաթղթերի ամբողջական մշակում՝ սկսած տեքստի և պատկերների առանձնացումից, մինչև տեքստի ամփոփում և դասակարգում՝ օգտագործելով ժամանակակից AI տեխնոլոգիաներ:

3.4. ՕԳՏԱԳՈՐԾՈՂԻ ԻՆՏԵՐՖԵՅՍ

Ինտերֆեյսի նախագծում

Օգտագործողի ինտերֆեյսը մշակվել է մատչելիության և հեշտ օգտագործման սկզբունքներով՝ օգտագործելով ժամանակակից վեբ դիզայնի լուծումներ: Ինտերֆեյսը նախագծվել է ջերմ, հարմարավետ գունային սխեմայով, որն օգտագործում է դեղին և ոսկեգույն երանգներ՝ ստեղծելով փաստաթղթի ընթերցման և մշակման համար հաճելի միջավայր: Համակարգի ամբողջական ինտերֆեյսը ցուցադրված է նկ. 3.1-ում:

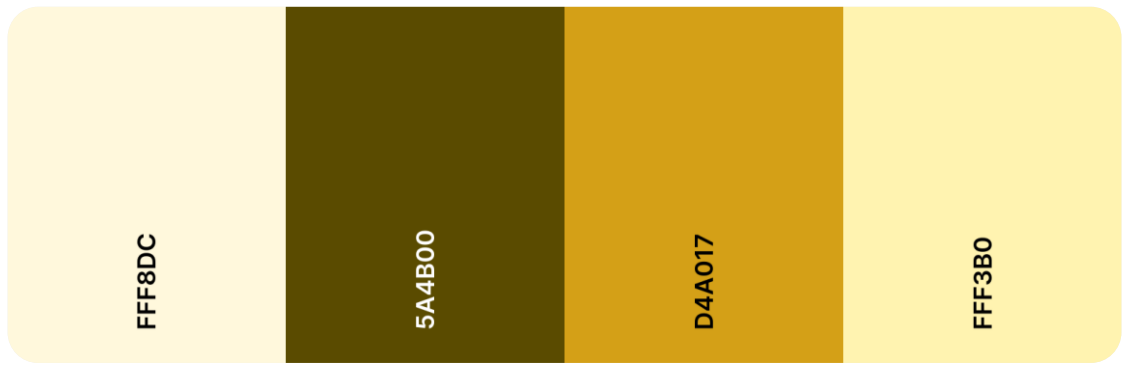
The screenshot displays a web interface for document processing. At the top right, there are language selection buttons for AM, EN, and RU. The main heading is 'Ամփոփման Գործիք' (Summary Tool). Below it, a paragraph explains that the tool converts PDF documents into a structured format, allowing users to upload a PDF file or a document image. The interface includes a section for 'Վերբեմնել PDF' (Upload PDF) with a dashed box for the file name and a checkbox for 'Միացնել Դասակարգումը' (Connect sorting). Below this is a 'Ամփոփման Արդյունք' (Summary Result) section with a large text area and buttons for 'Պատճենել' (Copy) and 'Ներբեմնել' (Paste). At the bottom, there is a 'Դասակարգման Արդյունք' (Sorting Result) section with a text area.

Նկ. 3.1. Օգտագործողի ինտերֆեյսի ընդհանուր տեսքը:

Գունային բազայի հիմքում ընկած են հետևյալ երանգները՝

- Ֆոնային գույն՝ #fff8dc (բաց, թղթային երանգ)
- Տեքստային գույն՝ #5a4b00 (մուգ շագանակագույն)
- Հիմնական ընդգծումներ՝ #d4a017 (ոսկեգույն)
- Կոնտեյներների ֆոն՝ #fff3b0 (բաց դեղին)

Ինտերֆեյսի գունային սխեման ցուցադրված է նկ. 3.2-ում:



Նկ. 3.2. Ինտերֆեյսի գունային սխեման:

Ինտերֆեյսի կառուցվածքը կազմված է հետևյալ հիմնական մասերից՝

- Լեզվի ընտրության հատված՝ վերևի աջ անկյունում (հայերեն, անգլերեն, ռուսերեն)
- Վերնագիր և համակարգի նկարագրություն
- Ֆայլի վերբեռնման հատված՝ ինտուիտիվ drag-and-drop ֆունկցիոնալությամբ
- Կարգավորումների հատված՝ դասակարգման ֆունկցիան միացնելու համար
- Ամփոփման գործողության կոճակ
- Արդյունքների հատված՝ ամփոփման և դասակարգման տեքստերի համար
- Արդյունքների կառավարման կոճակներ (պատճենել, ներբեռնել)

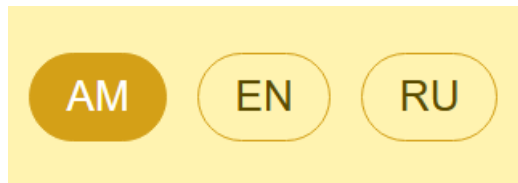
Համակարգի տպագրական ոճը հիմնված է Georgia տառատեսակի վրա, որն ապահովում է տեքստի լավ ընթերցելիություն: Համակարգի բաղադրիչները՝ կոճակները, ներբեռնման հատվածները և տեքստային դաշտերը, ունեն կլորացված անկյուններ (10-15px), որոնք բարելավում են օգտագործողի վիզուալ փորձառությունը:

Հիմնական ֆունկցիոնալություն

Համակարգն իրականացնում է հետևյալ հիմնական ֆունկցիաները՝

1. Բազմալեզու ինտերֆեյս

Համակարգը լիովին բազմալեզու է և աջակցում է երեք լեզու՝ հայերեն (AM), անգլերեն (EN) և ռուսերեն (RU): Օգտագործողը կարող է անցնել մի լեզվից մյուսին՝ օգտագործելով վերևի աջ անկյունում գտնվող լեզվի փոխարկիչները: Լեզվի փոփոխությունը դինամիկ է և անմիջապես թարմացնում է ինտերֆեյսի բոլոր տարրերը: Լեզվի ընտրության համակարգը ցուցադրված է նկ. 3.3-ում:



Նկ. 3.3. Բազմալեզու ինտերֆեյսի լեզվի ընտրության հատվածը:

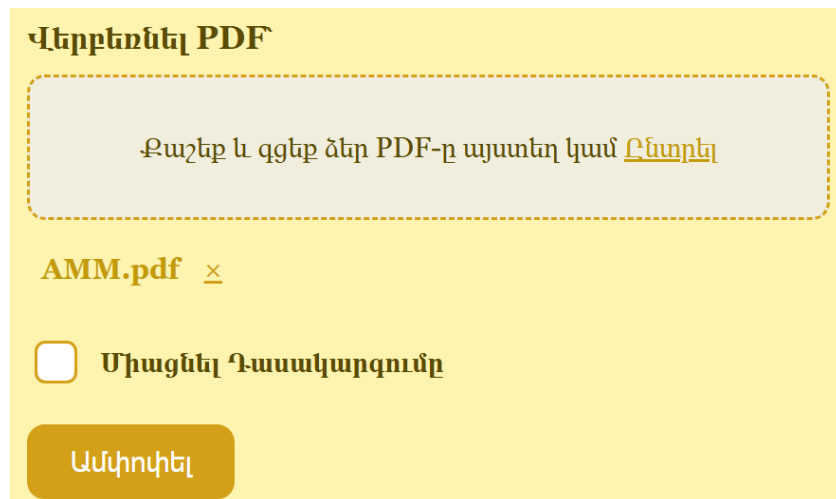
2. PDF ֆայլերի վերբեռնում

Համակարգն ապահովում է PDF ֆայլերի վերբեռնման երկու եղանակ՝

- Drag-and-drop՝ ֆայլը քաշելով դաշտի մեջ
- Ֆայլի ընտրիչ բրաուզերի միջոցով

Ֆայլի հաջող վերբեռնումից հետո համակարգը ցուցադրում է ֆայլի անունը և հնարավորություն է տալիս հեռացնել այն անհրաժեշտության դեպքում: Ֆայլի վերբեռնման տարածքը և վերբեռնված ֆայլի ցուցադրման եղանակը ներկայացված են նկ.

3.4-ում:

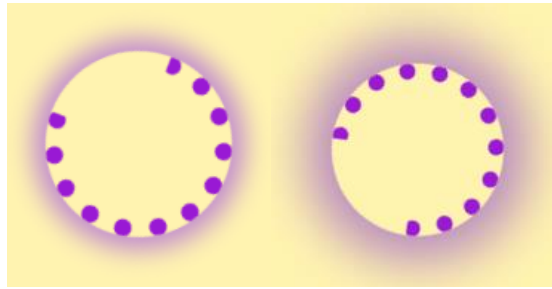


Նկ. 3.4. PDF ֆայլի վերբեռնման տարածքը և վերբեռնված ֆայլի ցուցադրումը:

3. Բովանդակության ամփոփման գործիք

Համակարգի հիմնական գործառնություն է PDF փաստաթղթերի մշակումը և դրանց բովանդակության ամփոփումը: Դա իրականացվում է արհեստական բանականության միջոցով, որը կարող է վերլուծել և ամփոփել փաստաթղթերը այն լեզվով, որով դրանք գրված են: Ամփոփման արդյունքները ցուցադրվում են առանձին դաշտում, որը հասանելի

է անմիջապէս վերբեռնումից հետո: Ամփոփման գործընթացի ընթացքում համակարգը
ցուցադրում է բեռնման ցուցիչ, որը տեսանելի է նկ. 3.5-ում:



Նկ. 3.5. Ամփոփման գործընթացի ընթացքում ցուցադրվող անիմացված բեռնման
ցուցիչը:

Ամփոփման արդյունքները ցուցադրվում են հստակ և կարդացվող ձևաչափով՝
հատուկ տարածքում, ինչպէս ցուցադրված է նկ. 3.6-ում:

Ամփոփման Արդյունք՝

Այս տեքստը ամփոփված ներկայացնում է արհեստական
բանականության (ԱԲ) տարբեր ասպեկտներ,
մասնավորապէս՝ փորձագիտական համակարգեր,
բազմազենոտ համակարգեր, գենետիկ ալգորիթմներ և ոչ
հստակ համակարգեր:

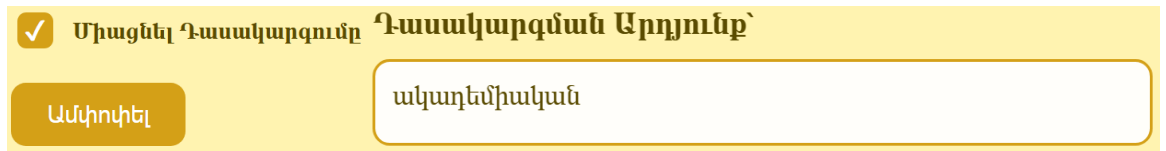
****Փորձագիտական համակարգերը**** նախատեսված են
փորձագետներին փոխարինելու և որոշումներ կայացնելու
համար: Դրանք լուծում են ախտորոշման,
կանխագուշակման, կառավարման և այլ խնդիրներ: Դրանց

Պատճենել
Ներբեռնել

Նկ. 3.6. Ամփոփման արդյունքների ցուցադրման տարածքը:

4. Բովանդակության դասակարգում

Համակարգն առաջարկում է լրացուցիչ ֆունկցիոնալություն՝ փաստաթղթի
բովանդակության դասակարգում ըստ տեսակի: Այս ֆունկցիան կարող է ակտիվացվել
օգտագործողի կողմից: Որոշում է փաստաթղթի տեսակը (օրինակ՝ անձնական,
ակադեմիական, և այլն): Դասակարգման ֆունկցիայի միացման դաշտը և արդյունքների
ցուցադրման տարածքը ներկայացված են նկ.3.7-ում:



Նկ. 3.7. Դասակարգման ֆունկցիայի կարգավորումը և արդյունքների ցուցադրումը:

5. Արդյունքների կառավարում

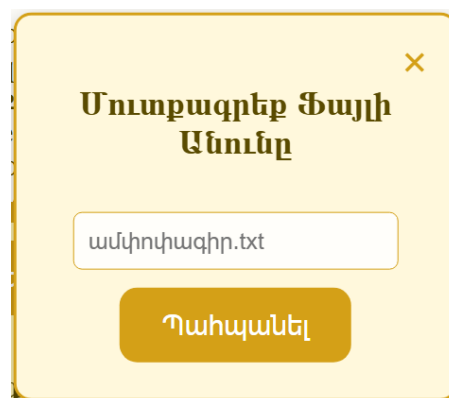
Համակարգը տրամադրում է արդյունքների հետ աշխատելու մի շարք գործիքներ՝

- Արդյունքների պատճենում սեղմատախտակին
- Արդյունքների ներբեռնում տեքստային ֆայլի տեսքով

Ներբեռնման ընթացքում համակարգը բացում է առանձին պատուհան, որտեղ հնարավոր է սահմանել ֆայլի անունը: Արդյունքների կառավարման կոճակները ցուցադրված են նկ. 3.8-ում, ֆայլի անվան մուտքագրման մոդալ պատուհանը՝ նկ. 3.9-ում:



Նկ. 3.8. Արդյունքների կառավարման կոճակները:



Նկ. 3.9. Ֆայլի անվան մուտքագրման մոդալ պատուհանը:

Օգտագործողի փորձի բարելավման մոտեցումներ

Օգտագործողի փորձի (UX) բարելավման համար համակարգում ներդրվել են մի շարք մոտեցումներ՝

1. Ինտուիտիվ և արձագանքող դիզայն

- Բոլոր կառավարման տարրերը հստակ տեսանելի են և պարզ նշված

- Կոճակների և փոխգործակցության տարրերի վրա կան վիզուալ արձագանքներ (hover, active վիճակներ)
- Համակարգը հարմարեցված է տարբեր սարքերի էկրանների չափերին

2. Պատասխանատու հետադարձ կապ

- Գործողությունների կատարման ընթացքում ցուցադրվում է պտտվող ծաղկանման ցուցիչ
- Օգտագործողին ցուցադրվում են տեղեկացնող հաղորդագրություններ (toast notifications)
- Ֆայլի հաջող վերբեռնման դեպքում ցուցադրվում է ֆայլի անունը

Հետադարձ կապի տարրերի օրինակ է "toast" ծանուցումը՝ ցուցադրված նկ. 3.10-ում:

Ամփոփումը պատճենվել է
սեղմատախտակին!

Նկ. 3.10. Toast ծանուցման օրինակը պատճենման գործողության ավարտից հետո:

3. Օգտագործողի սխալների կանխարգելում

- Համակարգը ստուգում է՝ արդյոք ֆայլ ընտրվել է նախքան ամփոփման գործընթացի սկսելը
- Սխալների դեպքում ցուցադրվում են հասկանալի հաղորդագրություններ
- Մոդալ պատուհանները ունեն փակելու հնարավորություն՝ հեշտ չեղարկման համար

4. Վիզուալ էսթետիկա և անիմացիաներ

- Բաղադրիչների սահուն անցումներ և ստվերներ՝ ընդգծելով տարածական հիերարխիան
- Մաքուր և կազմակերպված տեսակավորում՝ զերծ տեսողական ծանրաբեռնվածությունից
- Ներդրված անիմացիաներ՝ բարելավելու փոխգործակցությունը

5. Բազմալեզու աջակցություն

- Համակարգի բոլոր էլքային տվյալները (ամփոփումներ) գեներացվում են ներմուծված փաստաթղթի լեզվին համապատասխան

- Ինտերֆեյսը առաջարկում է լեզվի արագ փոխարկման հնարավորություն՝ հեշտացնելով տարբեր լեզվով խոսող օգտատերերի աշխատանքը։ Բազմալեզու համակարգի օրինակները՝ տարբեր լեզուներով ինտերֆեյսի ցուցադրմամբ, ներկայացված են նկ. 3.11-ում և նկ. 3.12-ում։

AM EN RU

Summarization Tool

This tool summarizes PDF files in Armenian, English, and Russian. The generated summary will be in the same language as the uploaded document.

Upload PDF:

Drag and drop your PDF here or [Browse](#)

☐ Enable Classification

Summarize

Summary Result:

Copy Download

Classification Result:

Նկ. 3.11. Բազմալեզու ինտերֆեյսի անգլերեն օրինակ:

AM EN RU

Инструмент Суммирования

Этот инструмент суммирует PDF файлы на армянском, английском и русском языках. Сгенерированный текст будет на том же языке, что и загруженный документ.

Загрузить PDF:

Перетащите сюда ваш PDF или [Выберите](#)

☐ Включить Классификацию

Суммировать

Результат Суммирования:

Копировать
Скачать

Результат Классификации:

Նկ. 3.12. Բազմալեզու ինտերֆեյսի ուսերեն օրինակ:

6. Մատչելիության լուծումներ

- Տեքստերն ունեն պատշաճ գունային կոնտրաստ ֆոնի հետ
- Ինտերակտիվ տարրերը հստակ տարբերակվում են ստատիկ տարրերից
- Օգտագործվել են համապատասխան մետադատաներ՝ պատշաճ էկրանի ընթերցման սարքերի աջակցության համար

Այս մոտեցումները համատեղ ապահովում են հարմարավետ և արդյունավետ օգտագործողի փորձառություն՝ PDF ֆայլերի ամփոփման համակարգում աշխատելիս: Համակարգը նախատեսված է որպես ինտուիտիվ և հեշտ օգտագործվող գործիք, որը կարող է ծառայել բազմազան լեզվական կարիքների ունեցող օգտատերերին:

3.5. API ԻՆՏԵՐՖԵՅՍ

API ինտերֆեյսի ընդհանուր նկարագրություն

API ինտերֆեյսը նախատեսված է համակարգի հիմնական ֆունկցիոնալությունը հասանելի դարձնելու երրորդ կողմի հավելվածների և ծառայությունների համար: Այն հնարավորություն է տալիս օգտագործողներին ինտեգրել ամփոփման և դասակարգման ֆունկցիոնալությունը իրենց սեփական համակարգերում՝ REST API ստանդարտի միջոցով:

API-ն իրականացնում է հետևյալ հիմնական գործառնությունները՝

- PDF փաստաթղթերի ներբեռնում և մշակում
- Ամփոփման գործողության իրականացում
- Պարամետրական դասակարգում՝ տարբեր լեզուներով
- API փաստաթղթավորման հասանելիություն

API հարցումների նկարագրություն

Համակարգն ունի երկու հիմնական API հարցում՝

1. **GET /api** - Վերադարձնում է API-ի փաստաթղթավորումը
2. **POST /api/summarize** - Իրականացնում է PDF փաստաթղթի ամփոփում և դասակարգում

GET /api հարցում

Այս հարցումը տրամադրում է API-ի ամբողջական փաստաթղթավորումը JSON ձևաչափով: Այն ներառում է բոլոր հասանելի հարցումների նկարագրությունները, պարամետրերը, պատասխանի կառուցվածքը և աջակցվող լեզուների ցանկը:

Պատասխանը ներառում է՝

- Բոլոր հասանելի վերջնակետերի ցանկը
- Յուրաքանչյուր վերջնակետի մեթոդները և նկարագրությունը
- Պահանջվող և ոչ պահանջվող պարամետրերի նկարագրությունը
- Պատասխանի կառուցվածքի նկարագրությունը
- Աջակցվող լեզուների ցանկ
- Դասակարգման հասանելի կատեգորիաների ցանկը բոլոր աջակցվող լեզուներով

Սա օգնում է մշակողներին ծանոթանալ API-ի հնարավորություններին և ճիշտ կիրառել այն:

POST /api/summarize հարցում

Սա համակարգի հիմնական հարցումն է, որը տրամադրում է PDF փաստաթղթերի ամփոփման և դասակարգման ֆունկցիոնալություն: Այն ընդունում է մուտքային տվյալները երկու ձևաչափով՝ multipart/form-data (ֆայլի ուղղակի վերբեռնման համար) և application/json (base64 կոդավորված PDF բովանդակության համար):

Մուտքային պարամետրեր

Multipart/form-data ձևաչափի պարամետրերը՝

- **file** (պարտադիր) - PDF ֆայլ
- **classification** (ոչ պարտադիր) - Boolean արժեք, որը ցույց է տալիս՝ արդյոք պետք է դասակարգել փաստաթուղթը (լռելային՝ false)
- **language** (ոչ պարտադիր) - Լեզվի կոդը դասակարգման կատեգորիաների համար (hy/en/ru) (լռելային՝ hy)

JSON ձևաչափի պարամետրերը՝

- **content** (պարտադիր) - PDF ֆայլի base64 կոդավորված բովանդակություն
- **filename** (ոչ պարտադիր) - Ֆայլի անունը
- **classification** (ոչ պարտադիր) - Boolean արժեք դասակարգման համար
- **language** (ոչ պարտադիր) - Լեզվի կոդը դասակարգման համար

Պատասխանի կառուցվածք

API վերադարձնում է JSON ձևաչափով պատասխան, որը ներառում է՝

- **summary** - Փաստաթղթի ամփոփված բովանդակությունը
- **classification** - Փաստաթղթի դասակարգումը (եթե պահանջվել է)
- **filename** - Բնօրինակ ֆայլի անունը
- **status** - Գործողության կարգավիճակը ("success" կամ "error")

Միայնի դեպքում API-ն վերադարձնում է համապատասխան սխալի կոդ և հաղորդագրություն:

API սահմանափակումներ և անվտանգություն

Համակարգում ներդրվել են մի շարք սահմանափակումներ և անվտանգության միջոցառումներ՝

- **Ֆայլի չափի սահմանափակում** - առավելագույնը 16ՄԲ
- **Ֆայլի տեսակի ստուգում** - ընդունվում են միայն PDF ֆայլեր
- **Ֆայլի անվան անվտանգություն** - ֆայլերի անվանումները անցնում են secure_filename միջոցով մաքրում
- **Ժամանակավոր ֆայլերի մաքրում** - վերբեռնված ֆայլերը հեռացվում են մշակումից հետո

API հուսալիության բարելավումներ

Համակարգում ներդրվել են մի շարք մեխանիզմներ API-ի հուսալիությունը բարելավելու համար՝

- **Սխալների մշակում** - բոլոր հնարավոր սխալները մշակվում են և վերադարձվում կանոնակարգված ձևաչափով
- **Ժամանակավոր ֆայլերի կառավարում** - օգտագործվում է tempfile մոդուլը ժամանակավոր ֆայլերի անվտանգ պահպանման համար
- **Ստուգումներ** - իրականացվում են ներմուծվող տվյալների հաստատման տարբեր ստուգումներ

API կիրառման հնարավորություններ

API ինտերֆեյսը բացում է բազմաթիվ կիրառության սցենարներ՝

1. **Ինտեգրացիա փաստաթղթերի կառավարման համակարգերում** - API-ն կարող է օգտագործվել գոյություն ունեցող փաստաթղթերի կառավարման համակարգերում՝ PDF փաստաթղթերի բովանդակության ավտոմատ ամփոփման համար:
2. **Վեբ հավելվածներում ընդգրկում** - API-ն կարող է կանչվել ցանկացած վեբ հավելվածից՝ փաստաթղթերի մշակման ֆունկցիոնալություն ավելացնելու համար:
3. **Բջջային հավելվածների աջակցություն** - Բջջային հավելվածները կարող են օգտագործել API-ն՝ փաստաթղթերի սկանավորման և մշակման համար:

4. **Ավտոմատացված բիզնես գործընթացներ** - API-ն կարող է լինել բիզնես գործընթացների ավտոմատացման մաս՝ փաստաթղթերի մեծ ծավալների մշակման համար:
5. **Բազմալեզու փաստաթղթերի մշակում** - API-ն աջակցում է տարբեր լեզուներով փաստաթղթերի մշակմանը և դասակարգմանը:

API ինտերֆեյսը զգալիորեն ընդլայնում է համակարգի կիրառման հնարավորությունները՝ թույլ տալով ինտեգրել այն տարբեր հարթակներում և կիրառել տարբեր սցենարներում: Այն նախագծված է՝ հաշվի առնելով RESTful ստանդարտները, անվտանգության պահանջները և ճկունությունը տարբեր կիրառություններում օգտագործելու համար:

Համակարգի API ինտերֆեյսը թույլ է տալիս մշակողներին և կազմակերպություններին ինտեգրել PDF փաստաթղթերի ավտոմատ ամփոփման և դասակարգման ֆունկցիոնալությունը իրենց սեփական լուծումներում՝ առանց վեր ինտերֆեյսի օգտագործման կարիքի, ինչը մեծապես բարձրացնում է համակարգի օգտակարությունը և կիրառելիությունը տարբեր բնագավառներում:

3.6. ՕԳՏԱԳՈՐԾՄԱՆ ՕՐԻՆԱԿՆԵՐ

Այս ենթաբաժնում ներկայացված են համակարգի օգտագործման գործնական օրինակներ՝ տարբեր լեզուներով և դասակարգումներով: Օրինակները ցույց են տալիս համակարգի բազմալեզու հնարավորությունները և ճկունությունը՝ տարբեր բովանդակության փաստաթղթերի մշակման համար:

Օրինակ 1: Հայերեն ինտերֆեյս և հայերեն բովանդակություն

Առաջին օրինակում ցուցադրվում է համակարգի աշխատանքը հայերեն ինտերֆեյսով և հայերեն բովանդակությամբ փաստաթղթի վրա: Փաստաթուղթը դասակարգվել է որպես «Ակադեմիական»:

Օրինակում օգտագործողը վերբեռնել է հայերեն լեզվով գրված գիտական փաստաթուղթ և ստացել է դրա ամփոփ տարբերակը: Համակարգը պահպանել է բովանդակության հիմնական գաղափարները՝ պահելով հայերեն լեզվի առանձնահատկությունները:

Նկար 3.13-ում ցուցադրված է հայերեն ինտերֆեյսով համակարգի աշխատանքի օրինակը և հայերեն բովանդակությամբ փաստաթղթի ամփոփման արդյունքը:

AM

EN

RU

Ամփոփման Գործիք

Այս գործիքը ամփոփում է PDF ֆայլերը հայերեն, անգլերեն և ռուսերեն լեզուներով: Ստեղծված ամփոփագիրը կլինի նույն լեզվով, ինչ վերբեռնված փաստաթուղթը:

Վերբեռնել PDF

Քաշեք և գցեք ձեր PDF-ը այստեղ կամ [Ընտրել](#)

Alg_miji.pdf

×

☒ Միացնել Դասակարգումը

Ամփոփել

Ամփոփման Արդյունք`

Ալգորիթմը հստակ սահմանված հաշվարկային ընթացակարգ է, որը մուտքային տվյալներից ստանում է որոշակի արդյունք: Ալգորիթմի բարդությունը ցույց է տալիս, թե ինչպես փոխվում է աշխատանքի ժամանակը մուտքային տվյալների չափի հետ: Ասիմպտոտիկ վերլուծությունը օգտագործում է Big-O, Omega և Theta նշումները՝ ալգորիթմի բարդությունը գնահատելու համար:

Այս տեքստում քննարկվում են հետևյալ ալգորիթմները.

- * ****Երկուական որոնում:**** Տեսակավորված զանգվածում տարր գտնելու արդյունավետ ալգորիթմ (ժամանակային բարդություն՝ $O(\log n)$).
- * ****Լերմուծման տեսակավորում:**** Պարզ տեսակավորման ալգորիթմ, որը յուրաքանչյուր տարրը տեղադրում է իր տեղում

Պատճենել

Լերբեռնել

Դասակարգման Արդյունք`

ակադեմիական

Նկ. 3.13. Հայերեն ինտերֆեյսով և հայերեն բովանդակությամբ փաստաթղթի ամփոփում:

Օրինակ 2: Անգլերեն ինտերֆեյս և ռուսերեն բովանդակություն

Երկրորդ օրինակում ցուցադրվում է համակարգի ճկունությունը՝ տարբեր լեզուներով աշխատելու հարցում: Օգտագործողը ընտրել է անգլերեն ինտերֆեյս, սակայն վերբեռնել է ռուսերեն լեզվով տեխնիկական փաստաթուղթ: Համակարգը հաջողությամբ ճանաչել է փաստաթղթի լեզուն և ամփոփումը տրամադրել ռուսերեն լեզվով, իսկ փաստաթուղթը դասակարգել որպես «Տեխնիկական»:

Այս օրինակը ցույց է տալիս համակարգի կարողությունը՝ հարմարվելու օգտագործողի նախընտրած ինտերֆեյսի լեզվին, միաժամանակ պահպանելով փաստաթղթի բնօրինակ լեզուն ամփոփման արդյունքներում:

Նկար 3.14-ում ցուցադրված է անգլերեն ինտերֆեյսով համակարգի աշխատանքի օրինակը և ռուսերեն բովանդակությամբ տեխնիկական փաստաթղթի ամփոփման արդյունքը:

The screenshot displays the 'Summarization Tool' interface. At the top right, there are language selection buttons for AM, EN, and RU. The main title 'Summarization Tool' is centered. Below it, a description states: 'This tool summarizes PDF files in Armenian, English, and Russian. The generated summary will be in the same language as the uploaded document.' The 'Upload PDF:' section features a dashed box with the text 'Drag and drop your PDF here or [Browse](#)'. Below this, a file named 'G_28_36.pdf' is shown with a close icon. A checkbox labeled 'Enable Classification' is checked. A 'Summarize' button is present. The 'Summary Result:' section shows a summary of a document titled '## Глава 1. Первое знакомство с организацией тестирования в Google - Краткое содержание'. The summary text is in Russian. Below the summary, there are 'Copy' and 'Download' buttons. The 'Classification Result:' section shows the word 'Technical' in a text box.

Նկ. 3.14. Անգլերեն ինտերֆեյսով և ռուսերեն բովանդակությամբ տեխնիկական փաստաթղթի ամփոփում:

Օրինակ 3: Ռուսերեն ինտերֆեյս և անգլերեն բովանդակություն

Երրորդ օրինակում ցուցադրվում է ռուսերեն ինտերֆեյսով համակարգի աշխատանքը անգլերեն բովանդակությամբ գովազդային բնույթի փաստաթղթի հետ: Փաստաթուղթը հաջողությամբ դասակարգվել է որպես «Գովազդային»:

Օրինակը ցույց է տալիս, թե ինչպես է համակարգը հաջողությամբ հարմարվում տարբեր լեզուների գուգակցմանը՝ պահպանելով բնօրինակ բովանդակության կարևոր առանձնահատկությունները:

Նկար 3.15-ում ցուցադրված է ռուսերեն ինտերֆեյսով համակարգի աշխատանքի օրինակը և անգլերեն բովանդակությամբ գովազդային փաստաթղթի ամփոփման արդյունքը:

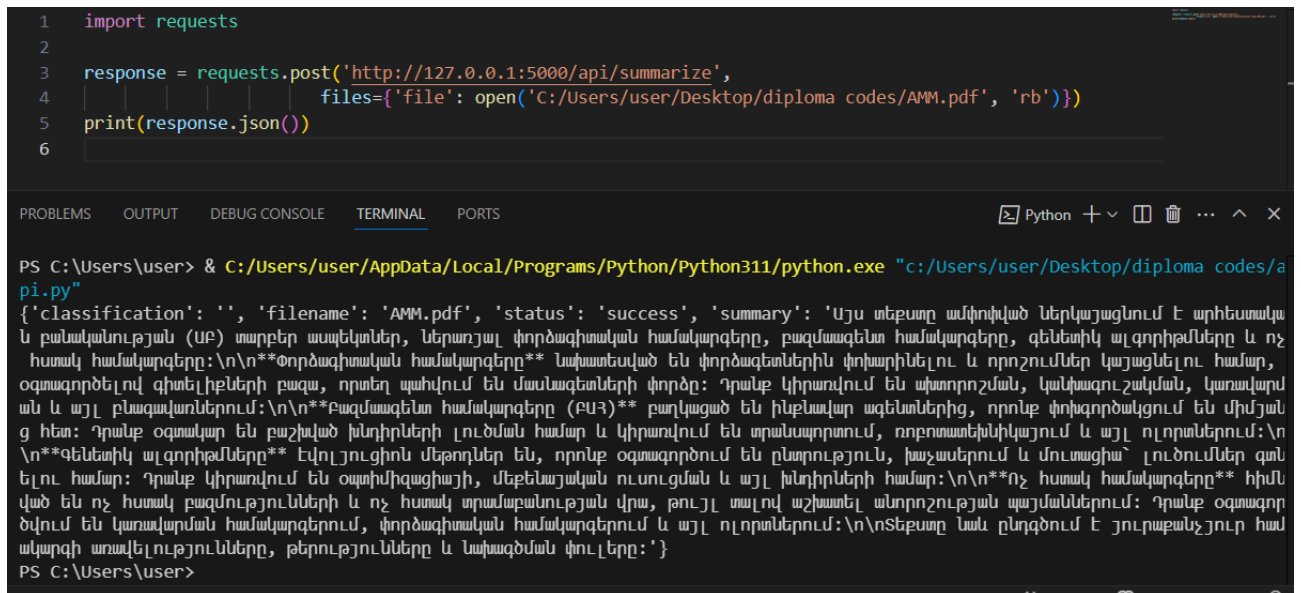
Նկ. 3.15. Ռուսերեն ինտերֆեյսով և անգլերեն բովանդակությամբ գովազդային փաստաթղթի ամփոփում:

Օրինակ 4: API-ի միջոցով կանչ

Չորրորդ օրինակում ցուցադրվում է համակարգի API ինտերֆեյսի օգտագործումը՝ հայերեն բովանդակությամբ փաստաթղթի մշակման համար: Ի տարբերություն նախորդ օրինակների, այստեղ օգտագործվում է ծրագրային ինտերֆեյսը՝ առանց վեբ միջերեսի:

API հարցման օրինակը՝

Նկար 3.16-ում ցուցադրված է API-ի միջոցով փաստաթղթի ամփոփման արդյունքը:



```
1 import requests
2
3 response = requests.post('http://127.0.0.1:5000/api/summarize',
4                           files={'file': open('C:/Users/user/Desktop/diploma codes/AMM.pdf', 'rb')})
5 print(response.json())
6
```

```
PS C:\Users\user> & C:/Users/user/AppData/Local/Programs/Python/Python311/python.exe "c:/Users/user/Desktop/diploma codes/a
pi.py"
{'classification': '', 'filename': 'AMM.pdf', 'status': 'success', 'summary': 'Այս տեքստը ամփոփված ներկայացնում է արհեստակա
ն բանականության (ԱԲ) տարբեր ապրելիսներ, ներառյալ փորձագիտական համակարգերը, բազմազգեստ համակարգերը, գենետիկ ալգորիթմները և ոչ
հստակ համակարգերը: Նախատեսված են փորձագետներին փոխարինելու և որոշումներ կայացնելու համար,
օգտագործելով գիտելիքների բազա, որտեղ պահվում են մասնագետների փորձը: Դրանք կիրառվում են փառորոշման, կանխագուշակման, կառավարմ
ան և այլ բնագավառներում: Նախատեսված են փորձագետները (ԲԱՅ)** բաղկացած են ինքնավար ազնուսներից, որոնք փոխգործակցում են միմյա
նից հետո: Դրանք օգտակար են բաշխված խնդիրների լուծման համար և կիրառվում են տրանսպորտում, ոռոգումային ինքնավարում և այլ ոլորտներում: Նա
խատեսված են փոխգործակցող ալգորիթմները** Էվոլյուցիոն մեթոդներ են, որոնք օգտագործում են ընտրության, խաչասերում և մուտացիա՝ լուծումներ գտն
ելու համար: Դրանք կիրառվում են օպտիմիզացիայի, մեքենայական ուսուցման և այլ խնդիրների համար: Նախատեսված են փոխգործակցող ալգորիթմները** Իմ
ված են ոչ հստակ բազմությունների և ոչ հստակ տրամաբանության վրա, թույլ տալով աշխատել անորոշության պայմաններում: Դրանք օգտագոր
ծվում են կառավարման համակարգերում, փորձագիտական համակարգերում և այլ ոլորտներում: Նախատեսված են ընդգծում և յուրաքանչյուր համ
ակարգի առավելությունները, թերությունները և նախագծման փուլերը: }
```

Նկ. 3.16. API-ի միջոցով հայերեն փաստաթղթի ամփոփման արդյունք:

Բերված օրինակները ցույց են տալիս համակարգի բազմալեզու հնարավորությունները և դրա կիրառման ճկունությունը տարբեր սցենարներում: Համակարգը հաջողությամբ մշակում է տարբեր լեզուներով գրված փաստաթղթեր և տրամադրում արդյունքները հարմարեցված ինտերֆեյսով կամ ծրագրային API-ի միջոցով:

ԳԼՈՒԽ 4. ԲՆԱՊԱՀՊԱՆՈՒԹՅՈՒՆ

4.1. ԷԼԵԿՏՐԱՄԱԳՆԻՍԱԿԱՆ ԴԱՇՏԻ ԱԶԴԵՑՈՒԹՅՈՒՆԸ ՇՐՋԱԿԱ ՄԻՋԱՎԱՅՐԻ ԵՎ ՄԱՐԴՈՒ ԱՌՈՂՋՈՒԹՅԱՆ ՎՐԱ

Էլեկտրական դաշտը ստեղծվում է լիցքակիր մասիկների կողմից: Մագնիսական դաշտը ստեղծվում է շարժվող լիցքակիրների կողմից: Էլեկտրական դաշտի մեծությունը բնութագրելու համար օգտագործվում է էլեկտրական դաշտի լարվածություն հասկացությունը, որը նշանակում են E տառով, չափման միավորն է Վ/մ-ը: Մագնիսական դաշտի մեծությունը բնութագրվում է մագնիսական դաշտի լարվածությամբ, որը նշանակում են H տառով, չափման միավորն է Ա/մ-ը: Ըստ նշանակության՝ էլեկտրամագնիսական դաշտը մատերիայի հիմնական ձև է: Էլեկտրամագնիսական դաշտի գոյության ֆիզիկական պատճառները պայմանավորված են նրանով, որ էլեկտրական դաշտի լարվածությունը գերազանցում է մագնիսական դաշտի լարվածությանը:

Էլեկտրական և մագնիսական դաշտերի լարվածությունները անընդհատ փոփոխվելով, գրգռում են մեկը մյուսին: Անշարժ կամ հավասարաչափ շարժվող լիցքավորված մասիկների կողմից առաջացած էլեկտրամագնիսական դաշտը անմիջապես կապված է այդ լիցքակիրների հետ, իսկ արագացմամբ շարժվող լիցքավորված մասնիկների կողմից առաջացած էլեկտրամագնիսական դաշտը «անջատվում» է նրանցից և գոյատևում է նրանցից անկախ էլեկտրամագնիսական ալիքների տեսքով: Էլեկտրամագնիսական ալիքները բնութագրվում են ալիքի երկարությամբ: Էլեկտրամագնիսական դաշտի կարևոր առանձնահատկությունը նրա բաժանումն է «մոտակա (սկզբնական)» և «հեռակա (վերջնական)» գոտիների: «Սկզբնական» գոտում էլեկտրամագնիսական դաշտը կարելի է համարել քվադրադոլական: Այս գոտին գտնվում է աղբյուրից ալիքի երկարությունից կարճ հեռավորությունների վրա: «Վերջնական» գոտին սկսվում է ալիքի երկարությունից 3 անգամ մեծ հեռավորություններից: Այն էլեկտրամագնիսական ալիքներից առաջացած գոտի է: Այս գոտում էլեկտրական և մագնիսական դաշտերի լարվածությունները միմյանց

հետ կապված են հետևյալ կերպ $E=377H$, որտեղ H -ը ալիքային դիմադրությունն է վակուումում, չափման միավորը Օհմ-ն է:

Անհատական համակարգիչներ

Համակարգիչներից օգտվողների առողջության վրա վատ ազդեցությունը հիմնականում կայանում է ինֆորմացիայի տեսողական արտացոլումը էլեկտրաճառագայթային խողովակի վրա: Մոնիտորի էկրանի էրգոնոմետրական պարամետրերն են. հակադարձ պատկերման փոքրացում՝ ինտենսիվ ներքին լուսավորման պայմաններում, հայելային արտացոլում մոնիտորի էկրանի առջևի մակերևույթից, պատկերի թարթման առկայություն՝ մոնիտորի էկրանի վրա:

Մոնիտորի էլեկտրամագնիսական դաշտը տատանվում է 20 Հց-ից մինչև 1 ԳՀց հաճախությամբ տիրույթում: Մոնիտորի էկրանին կա ստատիկ էլեկտրական լիցք: Մոնիտորը ունի ուլտրամանուշակագույն ճառագայթում 1050 նմ - 1 միլիմետր ալիքի երկարության տիրույթում: Ունի ռենտգենյան ճառագայթում:

Համակարգիչը որպես փոփոխական էլեկտրական դաշտի աղբյուր

Անհատական համակարգչի հիմնական բաղկացուցիչ մասեր հանդես են գալիս համակարգչային պրոցեսորը և տարբեր սարքեր՝ ստեղնաշար, պրինտեր, սկաներ և այլն: Յուրաքանչյուր անհատական համակարգիչ իր մեջ պարունակում է ինֆորմացիայի արտացոլման միջոցներ, որոնք կարող են կոչվել տարբեր ձևերով՝ մոնիտոր, դիսպլեյ: Որպես կանոն, նրա հիմքն է հանդիսանում էլեկտրոնաճառագայթային խողովակը:

Մոնիտորի աշխատանքի ժամանակ կինեսկոպի էկրանի վրա կուտակվում են էլեկտրաստատիկ լիցքեր, որոնք ստեղծում են էլեկտրաստատիկ դաշտ: Տարբեր հետազոտությունների արդյունքում և տարբեր պայմաններում չափվող էլեկտրաստատիկ դաշտերի լարվածությունները տատանվում են 8-75 կՎ/մ սահմաններում: Այս պայմաններում մոնիտորի հետ աշխատող մարդիկ ձեռք են բերում էլեկտրաստատիկ պոտենցիալ: Օգտվողների էլեկտրաստատիկ պոտենցիալը տատանվում է 1-5 կՎ միջակայքում: Էլեկտրաստատիկ դաշտի նկատելի ներդրում կատարում են ստեղնաշարը և մկնիկը:

Էլեկտրամագնիսական դաշտի ազդեցությունը մարդու առողջության վրա

Օրվա ընթացքում 2-6 ժամ մոնիտորի դիմաց աշխատողների նյարդային համակարգի խախտում կատարվում է 4.6 անգամ ավելի հաճախ, քան վերահսկողական խմբերում, սրտանոթային համակարգի հիվանդություն՝ 2 անգամ ավելի հաճախ: Շնչառական ուղիների հիվանդություններ ձեռք են բերում 1.9 անգամ հաճախ, իսկ հենաշարժական համակարգի հիվանդություններ՝ 3.1 անգամ: Համակարգչից օգտվողի ֆունկցիոնալ վիճակի հետազոտումը ցույց է տվել, որ նույնիսկ շատ ավելի քիչ ժամանակով (45 րոպե) օգտվողների օրգանիզմում, մոնիտորի էլեկտրամագնիսական ճառագայթների ազդեցությամբ, տեղի են ունենում հարմոնիկ վիճակի և ուղեղի կենսահոսքերի յուրահատուկ փոփոխություններ: Այս երևույթը ավելի հաճախ լինում է կանանց մոտ:

Աչքերը փառակալվում են, այսինքն աչքերի դիմաց շղարշի շերտ է պատում, աչքերը հոգնում են, հիվանդ են դառնում: Առաջանում են գլխացավեր, քունը խանգարվում է, փոփոխվում է օրգանիզմի հոգեֆիզիկական վիճակը: Սթրես հաճախ առաջանում է դիսփլեյից օգտվողների մոտ: Հայտնաբերված է, որ սթրեսի աղբյուր կարող են հանդիսանալ՝ գործունեության տեսքը, համակարգչի բնութագրիչ յուրահատկությունները, օգտագործվող ծրագրային ապահովումը, աշխատանքի կազմակերպումը, սոցիալական տեսակետները: Մարդու սթրեսային վիճակում մասլը կարող է բերել քնի խանգարման, սրտի զարկի հաճախականության փոփոխման: Որպես պաշտպանության հիմնական միջոց առաջարկում են պաշտպանիչ գոտիներ մոնիտորի էկրանի համար: Դրանք օգտագործվում են սպառողի վրա փաստակար գործոնների ազդեցության դեմ պայքարի համար, լավացնում են էկրանի՝ մոնիտորի պարամետրերը, փոքրացնում են մոնիտորի ճառագայթումը, որն ուղղված է դեպի սպառողը: Առաջին հերթին մարդու նյարդային համակարգը շատ զգայուն է էլեկտրամագնիսական դաշտի նկատմամբ, երկրորդ հերթին՝ էլեկտրամագնիսական դաշտը ունենում է ինֆորմացիոն ազդեցություն, որն ազդում է մարդու վրա ջերմային էֆեկտի սահմանային էֆեկտից ցածր մակարդակների դեպքում: Տարբեր երկրների հետազոտողների փորձարարական տվյալները վկայում են էլեկտրամագնիսական դաշտի կենսաբանական բարձր ակտիվության մասին:

Ժամանակակից տեսության տեսանկյունից էլեկտրամագնիսական դաշտի ճառագայթման հարաբերական բարձր մակարդակների դեպքում տեղի է ունենում ջերմային ազդեցություն: Իսկ էլեկտրամագնիսական դաշտի համեմատաբար ցածր մակարդակների ժամանակ ազդեցությունը օրգանիզմի վրա ոչ ջերմային կամ ինֆորմացիոն բնույթի է: Էլեկտրամագնիսական դաշտի կենսաբանական շրջանի ազդեցության մանրամասն հետազոտությունները ցույց են տալիս, որ օրգանիզմի նյարդային, իմունային, էնդոկրին և սեռական համակարգերը առավել զգայուն են այդ դաշտերի նկատմամբ: Օրգանիզմի այդ դաշտերը համարվում են կրիտիկական: Այս համակարգերի ռեակտիվության պետք է հաշվի առնվի էլեկտրամագնիսական դաշտի ազդեցության դեպքում: Երկարատև ազդեցության դեպքում էլեկտրամագնիսական դաշտի կենսաբանական էֆեկտը կուտակվում է, որի արդյունքում հնարավոր է առանձին հետևանքների զարգացում, օրինակ՝ կենտրոնական նյարդային համակարգի դեգեներատիվ պրոցես, արյան քաղցկեղ (լեյկոզ), ուղեղի ուռուցքներ, հորմոնային հիվանդություններ: Գերբարձր հաճախականության էլեկտրամագնիսական դաշտի (ԳԲՀ ԷՄԴ) ազդեցությունը կենդանի օրգանիզմների վրա վաղուց է գրավում հետազոտողների և գործնական առողջապահության մասնագետների ուշադրությունը իր հավանական, սակայն ոչ բավարար ուսումնասիրված տվյալներով: Հայտնի է, որ արդյունաբերության, զինության բնագավառներում և զինվորական դործում միկրոալիքների լայն կիրառման հետևանքով աճում է այն անձանց թիվը, ովքեր մասնակցում են ԳԲՀ ԷՄԴ գեներատորների գործարանային պատրաստմանը, փորձարկմանը, շահագործմանն ու վերանորոգմանը: ԳԲՀ ԷՄԴ դաշտը նյութի հատուկ տեսակ է: Տարբերվում են ռադիոալիքներ (բարձր հաճախականության ԳԲՀ, ուլտրաբարձր հաճախականության ՈՒԲՀ, գերբարձր հաճախականության ԳԲՀ), ինֆրակարմիր, տեսանելի (լույս), ուլտրամանուշակագույն (ՈՒՄ) ճառագայթներ, ռենտգենյան և գամմա ճառագայթներ: Այս բաժանման հիմնական ֆիզիկական չափորոշիչը ալիքի երկարությունն ու հաճախականությունն են: Համարվում է, որ ԳԲՀ ԷՄ կամ միկրոալիքների կարևոր չափորոշիչ է դրա ինտենսիվությունը, այսինքն էներգետիկ բնութագիրը: Այն գնահատվում է էներգիայի հոսքի խտության (ԷՀԽ) մեծությամբ ճառագայթման էներգիայով, որն անցնում է ալիքների տարածման ուղղահայաց 1 սմ² մակերեսով: Չափման միավորներն են Վտ/սմ² կամ ՄՎտ/սմ²:

ԳԲՀ ԷՄ դաշտի կենսաբանական ազդեցության ուսումնասիրությունը հանգեցրել է այն եզրակացության, որ ճառագայթահարվող օբյեկտի մակերևույթին ընկնող ամբողջ էներգիան չէ, որ ներգործում է, այլ այն մասը, որը կլանվում է տվյալ օբյեկտի կողմից: Ցույց է տրվել, որ միկրոալիքների էներգիայի ընդհանուր քանակից սովորաբար կլանվում է 50 տոկոսից ոչ ավելին, մնացած մասն անդրադառնում է: Ընդ որում կլանման ու ներթափանցման մեծությունը որոշվում է հյուսվածքի բնույթով. դիէլեկտրիկ հատկություններով և ԳԲՀ ԷՄ դաշտի ալիքի երկարությամբ: Ներթափանցման խորությունը մոտավորապես հավասար է ալիքի երկարության 1/10 մասին:

Տարբերում են միկրոալիքների ջերմային (ԷՀԽ-ը գերազանցում է 10 ՄՎտ/սմ²) և ոչ ջերմային կամ յուրահատուկ (ԷՀԽ-ը ավելի քիչ է քան 10 ՄՎտ/սմ²) ազդեցություն: Ջերմային ազդեցության օրինակ է օրգանիզմի առանձին մասերի և հանգույցների ջերմաստիճանի բարձրացումը, իսկ ոչ ջերմային՝ հյուսվածքների միկրոպրոցեսների փոփոխությունը, սպիտակուցային փոխանակության խախտումը:

ԳԼՈՒԽ 5. ԿԵՆՍԱԳՈՐԾՈՒՆԵՈՒԹՅԱՆ ԱՆՎՏԱՆԳՈՒԹՅՈՒՆ

5.1. ԷԼԵԿՏՐՈՆԱՅԻՆ ՓԱՍՏԱԹՂԹԵՐԻ ԶԳԱՅՈՒՆ

ՏԵՂԵԿԱՏՎՈՒԹՅԱՆ ԱՆՀԱՍԱՆԵԼԻՈՒԹՅՈՒՆ

Արդի դարաշրջանում համակարգիչն անփոխարինելի գործիք է դարձել մարդու առօրյայում՝ օգտագործվելով թե՛ տանը, թե՛ գրասենյակում, թե՛ խանութում, թե՛ արտադրական ոլորտում, և անգամ ամենօրյա կենցաղում: Այլ խոսքով՝ առանց դրա դժվար է պատկերացնել ժամանակակից կյանքը:

Համակարգիչները հիմնականում կիրառվում են որպես տեղեկատվության մշակման միջոց, և դրանց ներդրումը արմատապես փոխել է ոչ միայն գրասենյակային աշխատանքի բնույթը, այլև աշխատանքի կազմակերպման սկզբունքներն ու անվտանգության պահանջները: Աշխատանքային անվտանգության կանոնների խախտումը կարող է բացասաբար ազդել աշխատակցի առողջության վրա՝ հանգեցնելով ֆիզիկական խնդիրների, ինչպես նաև հոգեբանական լարվածության, տագնապի և սթրեսի:

Բացի այդ, անձնական համակարգիչներում պահվող էլեկտրոնային փաստաթղթերը կարող են ներառել զգայուն տեղեկատվություն՝ անձնական, ֆինանսական, աշխատանքային և այլ բնույթի: Տեղեկատվության հասանելիությունը կարող է վտանգվել տարբեր գործոնների՝ տեխնիկական անսարքությունների կամ չարամիտ հարձակումների պատճառով, ինչը կարող է լուրջ հետևանքներ ունենալ ինչպես անհատների, այնպես էլ կազմակերպությունների համար:

Կիբեռանվտանգության կարևորությունը

Կիբեռանվտանգությունը ներկայացնում է համակարգիչների, ցանցերի, ծրագրային ապահովման և տվյալների պաշտպանությունը չարտոնված մուտքից կամ չարամիտ գործողություններից: Կիբեռանվտանգությունը կարևոր է անձնական և կազմակերպական տվյալների պաշտպանվածությունը ապահովելու համար: Ժամանակակից աշխարհում տեղեկատվական համակարգերը հաճախ են ենթարկվում չարամիտ հարձակումների, որոնց հետևանքները կարող են լինել՝

1. Ֆինանսական վնասներ

- Գողացված բանկային տվյալները կարող են օգտագործվել հանցավոր նպատակներով:

2. Անձնական տվյալների բացահայտում

- Հակերները կարող են հասանելի դառնալ անձնական լուսանկարների, փաստաթղթերի կամ այլ զգայուն տեղեկատվությանը:

3. Կարողությունների կորուստ

- Համակարգիչը կարող է ենթարկվել ransomware (գումար պահանջող վիրուս) հարձակման, ինչը կհանգեցնի տվյալների ժամանակավոր կամ մշտական անհասանելիության:

Կիրեռանվտանգության բարձր մակարդակի ապահովումը թույլ է տալիս՝

1. Պաշտպանել անձնական տվյալները

- Կանխարգելելով գաղտնաբառերի կամ անձնական փաստաթղթերի գողությունը:

2. Նվազեցնել ֆինանսական ռիսկերը

- Ֆինանսական տվյալների պաշտպանության բացակայությունը կարող է հանգեցնել հաշվեհամարների թալանի կամ վնասված բանկային գործարքների:

3. Համակարգերի անխափան աշխատանք

- Հակերային հարձակումները հաճախ դառնում են ծրագրային խափանումների պատճառ, ինչը կարող է խանգարել աշխատանքային գործընթացներին:

Տեղեկատվության արտահոսքի հիմնական պատճառները

1. Զարամիտ ծրագրեր (Malware)

- Վիրուսներ, trojans և spyware ծրագրեր, որոնք կարող են տեղադրվել համակարգչում օգտատիրոջ անուշադրության հետևանքով:

2. Ֆիշինգ հարձակումներ (Phishing)

- Էլ. փոստով կամ սոցիալական ցանցերով ստացած հաղորդագրություններ, որոնք կեղծելով վստահելի աղբյուրներ՝ փորձում են գողանալ գաղտնաբառերը կամ այլ ինֆորմացիա:

3. Wi-Fi ցանցերի ոչ պաշտպանվածությունը

- Անապահով հանրային Wi-Fi ցանցերը, ինչպես նաև անձնական Wi-Fi ցանցի ոչ պատշաճ ապահովումը կարող են հեշտացնել տվյալների «գողությունը»:

4. Համակարգչի ֆիզիկական գողություն

- Համակարգիչը կորցնելիս կամ գողանալիս տվյալները կարող են ընկնել անհայտ անձանց ձեռքը:

Հնարավոր բացասական հետևանքներ

1. Համակարգի աշխատանքի խափանում

- Տեխնիկական խնդիրները կամ վիրուսային հարձակումները կարող են հանգեցնել տվյալների կորստի կամ համակարգչի անաշխատունակության:

2. Հավատարմության կորուստ

- Կազմակերպությունների համար տվյալների արտահոսքը հաճախ հանգեցնում է հաճախորդների վստահության կորստի:

3. Օրենքի հետ կապված խնդիրներ

- Անձնական տվյալների արտահոսքը կարող է հակասել տվյալների պաշտպանության միջազգային օրենքներին (օրինակ՝ GDPR):

4. Վարկանիշի և վստահության կորուստ

- Կազմակերպությունների համար տվյալների արտահոսքը հաճախ հանգեցնում է հաճախորդների և գործընկերների վստահության կորստի:

Կանխարգելման միջոցառումներ

1. Գաղտնաբառերի ուժեղացում

- Օգտագործել բարդ գաղտնաբառեր և դրանք պարբերաբար փոխել:

2. Անտիվիրուսային ծրագրերի օգտագործում

- Տեղադրել հուսալի անվտանգության ծրագրեր՝ չարամիտ ծրագրերից պաշտպանվելու համար:

3. Տվյալների պահուստավորում

- Պարբերաբար պահուստավորել կարևոր տվյալները արտաքին կրիչների կամ ամպային ծառայությունների միջոցով:

4. Կրթություն և իրազեկում

- Ծանոթանալ կիրառական տեխնոլոգիաների հիմնական մեթոդներին և սովորել ճանաչել կեղծ հաղորդագրությունները:

5. Ֆայլերի գաղտնագրում

- Օգտագործել ֆայլերի կոդավորման գործիքներ՝ տվյալները չարտոնված անձանց համար անհասանելի դարձնելու համար:

Էլեկտրոնային փաստաթղթերի տեղեկատվության հասանելիությունն ու պաշտպանությունը ժամանակակից թվային միջավայրում կարևոր են յուրաքանչյուրի համար: Կիրառական տեխնոլոգիաների պատշաճ ապահովումը ոչ միայն կկանխարգելի տվյալների կորուստը, այլև կպաշտպանի ձեր անձնական և ֆինանսական անվտանգությունը: Տեղեկատվության պաշտպանության լավագույն միջոցը հնարավոր ռիսկերի մասին իրազեկվածությունն է և պարբերական անվտանգության միջոցառումների իրականացումը: Առանց կիրառական տեխնոլոգիաների, տվյալների կորուստն ու փաստաթղթի ռիսկերը կարող են դառնալ իրական վտանգ:

ԳԼՈՒԽ 6. ՏՆՏԵՍԱԳԻՏՈՒԹՅՈՒՆ

6.1. ԷԼԵԿՏՐՈՆԱՅԻՆ ՓԱՍՏԱԹՂԹԵՐԻ (PDF) ԲՈՎԱՆԴԱԿՈՒԹՅԱՆ ՃԱՆԱԶՄԱՆ ԵՎ ԱՄՓՈՓՄԱՆ ԱՎՏՈՄԱՏԱՑՎԱԾ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՄԱՆ ԻՆՔՆԱՐԺԵՔԸ

Ժամանակակից տնտեսական պայմաններում ցանկացած գործունեություն առնչվում է տնտեսական հարցերի հետ: Նոր տեխնիկայի, տեխնոլոգիական պրոցեսների, տեխնիկա-տնտեսական հիմնավորման համար, մասնավորապես մեր աշխատանքի պայմաններում կարևորագույն տնտեսական հիմնահարցերից է ինքնարժեքի որոշումը:

Արտադրանքի կամ ծառայությունների ինքնարժեքը՝ դա արտադրանքի (ծառայությունների) արտադրության և իրացման վրա կատարված բոլոր ծախսերի գումարն է դրամական արտահայտությամբ:

Ինքնարժեքի մեջ իրենց արտահայտությունն են գտնում սպառված շրջանառու ֆոնդերը, կենդանի աշխատանքի մի մասը, որը աշխատողներին վճարում է աշխատավարձի ձևով:

Ինքնարժեքի մեջ մտնող ծախսերը դասակարգվում են ըստ տնտեսական տարերի և ըստ կալկուլյացիոն հոդվածների:

Ներկայումս կիրառվում է ծախսերի ըստ կալկուլյացիոն հիմնական հոդվածների հետևյալ դասակարգումը՝

1. համալրող առարկաներ,
2. էլեկտրաէներգիայի ծախսեր,
3. աշխատողների հիմնական աշխատավարձ,
4. աշխատողների լրացուցիչ աշխատավարձ,
5. սարքավորումների շահագործման և պահպանման ծախսեր,
6. տարածքի վարձակալության համար ծախս,
7. ընդհանուր տնտեսական ծախսեր:

Լրիվ ինքնարժեք (1-7 կետերի գումարը):

Փաթեթը, որի ինքնարժեքն ու գինը ենթակա է որոշման, իրենից ներկայացնում է՝ Էլեկտրոնային փաստաթղթերի (PDF) բովանդակության ճանաչման և ամփոփման ավտոմատացված համակարգ:

Հաշվարկի համար ելքային տվյալներ են հանդիսանում՝

- փաթեթի մեջ մտնող համալրող առարկաների քանակն ու անվանացանկը,
- ժամանակի ամփոփ նորմերը, աշխատանքի կարգն ու աշխատավարձի ձևերը,
- ժամավճարային և գործարքային պարգևատրման չափերը (26 %),
- լրացուցիչ աշխատավարձի չափերը (15 %),
- սարքավորումների պահպանման և շահագործման ծախսերի դրույքաչափերը,
- ընդհանուր տնտեսական ծախսերի դրույքաչափը (129 %):

1. Համալրող առարկաների ծախսի հաշվարկ

Գնված բաղադրիչների արժեքը որոշվում է հետևյալ բանաձևով՝

$$\Sigma_{կֆհա} = \Sigma \Phi_i * Q_i,$$

որտեղ Φ_i -ն i-րդ տեսակի գնված բաղադրիչների քանակն է, հատ, Q_i -ն՝ i-րդ գնված բաղադրիչի միավորի գինը: Հաշվարկի արդյունքները բերված են աղ. 6.1-ում:

Աղյուսակ 6.1.

Համալրող առարկաների ծախսի հաշվարկի արդյունքներ:

Բաղադրիչի անվանումն ու տեսակը	Մեկ փաթեթին ընկնող քանակ.հատ	Միավորի գինը.դրամ	Մեկ փաթեթին ընկնող արժեք.դրամ
Արագամաշ առարկաներ	8	250000	2000000
Գրասենյակային տարրեր /ֆլեշ կրիչ, ներկանյութ և այլն/:	3	15000	45000
Ընդամենը			2045000

Ընդամենը՝ 2045000 դրամ:

2. Էլեկտրաէներգիայի ծախսի հաշվարկ

Համակարգիչները և կապի ու SS այլ սարքավորումները աշխատեցնելու համար Էլեկտրաէներգիայի տարեկան ծախսը որոշվում է հետևյալ բանաձևով՝

$$E = \Phi * U_{\text{է}},$$

որտեղ Φ -ն Էլեկտրաէներգիայի տարեկան ծախսն է, $U_{\text{է}}$ -ն՝ 1կվ/ժ Էլեկտրաէներգիայի արժեքը (50 դրամ, ըստ գործարանային գների):

$$\Phi = Z_{\text{տհ}} \Phi_{\text{հ}} Q_{\text{բ}} \gamma_{\text{մ}} / \gamma_1 \gamma_2,$$

որտեղ $Z_{\text{տհ}}$ -ն հաստոցների Էլեկտրաշարժիչների ընդհանուր հզորությունն է՝ $Z = 1$ կՎտ, $\Phi_{\text{հ}}$ -ն՝ հաստոցների աշխատաժամանակի տարեկան իրական ֆոնդը՝ 25 օր, $Q_{\text{բ}}$ -ն՝ հաստոցների բեռնվածության միջին գործակիցը՝ 0.9, $\gamma_{\text{մ}}$ -ն՝ կորուստներն են ցանցում՝ 0.8, γ_1 -ը՝ միաժամանակ աշխատող հաստոցների շարժիչի Օ.Գ.Գ.-ն՝ 0.95, γ_2 -ը՝ շարժիչի Օ.Գ.Գ.-ն՝ 0.9, $Q_{\text{բ}} = 0.9$, $\gamma_{\text{մ}} = 0.8$, $\gamma_1 = 0.95$, $\gamma_2 = 0.9$:

Մեր օրինակի համար՝

$\Phi = 1000 / 0.95 * 0.9 = 1169,5$ կՎտ/ժ, $E = 1169,5 * 50 \approx 50000$ դրամ/ամսական և $50000 * 12 = 600000$ դրամ/տարեկան:

3. Աշխատողների հիմնական աշխատավարձի հաշվարկը

Աշխատողների հիմնական աշխատավարձի մեջ մտնում են՝

- գործարքային դրույքաչափերով աշխատավարձ,
- ժամավճարային աշխատավարձ,
- պարգևավճար:

Գործարքային աշխատավարձն ըստ տարիֆային համակարգի որոշվում է հետևյալ բանաձևով՝

$$U_{\text{հիմ.}} = \sigma_{\eta} * U_{\text{արտ.}},$$

որտեղ՝ σ_{η} -ն ժամային դրույքաչափն է, $U_{\text{արտ.}}$ -ն՝ ժամային նորմը: Հաշվարկման արդյունքները բերված են աղ. 6.2-ում:

Աշխատողների հիմնական աշխատավարձի հաշվարկի արդյունքներ:

Գործառույթի հաջորդական.	Վճարման ձև	Ժամանա- կային նորմ	Ժամային դրույք	Տարիֆային ֆոնդ
Նախապատ- րաստում	Գործարքա- պարգևա- վճարային	30	3500	105000
Մշակում	Գործարքա- պարգևա- վճարային	26	5000	130000
Կարգավորում	Գործարքա- պարգևա- վճարային	30	5000	150000
Տեստավորում	Գործարքա- պարգևա- վճարային	32	4500	144000
Ընդամենը				529000

Պարգևատրման չափը որոշվում է հետևյալ բանաձևով՝

$$\Pi = U_{\text{հիմ}} * \Pi_{\text{դ}} / 100\%,$$

որտեղ՝ $\Pi_{\text{դ}}$ - պարգևատրման դրույքաչափ, %

$$\Pi = 529000 * 26 / 100 = 137540 \text{ դրամ:}$$

Ընդամենը հիմնական աշխատավարձը կկազմի՝

$$137540 + 529000 = 666540 \text{ դրամ/ամսական կամ՝}$$

$$666540 * 12 = 7998480 \text{ դրամ/տարեկան:}$$

4. Աշխատողների լրացուցիչ աշխատավարձի հաշվարկը

Լրացուցիչ աշխատավարձի մեջ մտնում են՝ հերթական և լրացուցիչ գործողությունների, արձակուրդների վճարները, պետական հանձնարարականների կատարման հետ կապված ծախսերը և այլն: Աշխատողների լրացուցիչ աշխատավարձը հաշվարկվում է հետևյալ բանաձևով՝

$$U_{\text{լր.}} = \text{Ը}U_{\text{հիմ.}} * U_{\text{լր.դ.}}/100,$$

որտեղ՝ $\text{Ը}U_{\text{լր.դ.}}$ – Ընդհանուր հիմնական աշխատավարձն է, իսկ $U_{\text{լր.դ.}}$ -ն լրացուցիչ աշխատավարձի դրույքաչափն է, %:

$$U_{\text{լր.}} = 529000 * 15 / 100 = 79350 \text{ դրամ/ամսական կամ}$$

$$79350 * 12 = 952200 \text{ դրամ/տարեկան:}$$

5. Սարքավորումների պահպանման և շահագործման ծախսերի հաշվարկը

Սարքավորումների պահպանման և շահագործման ծախսերի թվին են պատկանում ամորտիզացիոն, ընթացիկ վերանորոգման, տրանզիտորային միջոցների, գործիքների և հարմարանքների վերանորոգման և այլ ծախսերը: Սարքավորումների պահպանման և շահագործման ծախսերի թվին են պատկանում ամորտիզացիոն, ընթացիկ վերանորոգման, տրանզիտորային միջոցների, գործիքների և հարմարանքների վերանորոգման և այլ ծախսերը:

Հիմնական միջոցների տարեկան ամորտիզացիան (U_s) հաշվարկվում է հետևյալ բանաձևով՝

$$U_s = Z_u / \text{Ն},$$

որտեղ Z_u -ն հիմնական միջոցների սկզբնական արժեքն է, Ն-ն հիմնական միջոցների օգտակար գործունեության ժամկետն է: Հաշվարկման արդյունքները բերված են աղ. 6.3-ում:

Սարքավորումների պահպանման և շահագործման ծախսերի հաշվարկի արդյունքներ:

Հիմնական միջոցի անվանումն ու տեսակը	Հիմնական միջոցի սկզբնական արժեքը	Ամորտիզացիոն հատկացումներ	
		ՀՄ օգտակար գործողության ժամկետ, տարի	Ամորտիզացիոն ծախս
Սարքավորումներ, շենքեր, այլ հիմն. միջոցներ	48750000	5	9750000
Ընդամենը			9750000

Սարքավորումների պահպանման և շահագործման ծախսերի գումարը բերված է աղ. 6.4-ում և որոշվում է հետևյալ կերպ՝

$$\bar{\sigma}_{\text{պ.շ.}} = U_{\text{հիմ}} * 12 * \bar{\sigma}_{\text{պ.շ.դ.}} / 100,$$

որտեղ՝ $\bar{\sigma}_{\text{պ.շ.դ.}}$ - սարքավորումների պահպանման և շահագործման ծախսերի դրույքաչափն է, $U_{\text{հիմ}}$ - աշխատողների տարեկան հիմնական աշխատավարձը:

$$\bar{\sigma}_{\text{պ.շ.}} = 7998480 * 2.1 / 100 = 167968.08 \quad \text{դրամ:}$$

Սարքավորումների ընթացիկ վերանորոգման ծախսը որոշվում է հետևյալ բանաձևով՝

$$\bar{\sigma}_{\text{ը.վ.}} = U_{\text{հիմ}} * 12 * \bar{\sigma}_{\text{ը.վ.դ.}} / 100,$$

որտեղ՝ $\bar{\sigma}_{\text{ը.վ.դ.}}$ - սարքավորումների ընթացիկ վերանորոգման դրույքաչափն է:

$$\bar{\sigma}_{\text{ը.վ.}} = 7998480 * 4.5 / 100 = 359931.6 \quad \text{դրամ:}$$

Սարքավորումների ընթացիկ վերանորոգման ծախսի արդյունքներ:

Ծախսի անվանումը	Դրույքաչափը	Տարեկան ծախսը
Սարքավորումների պահպանման և շահագործման ծախսեր	2.1	167968.08
Սարքավորումների ընթացիկ վերանորոգում	4.5	359931.6
Սարքավորումների ամորտիզացիա	-	9750000
Ընդամենը		10277899.68

6. Տարածքի վարձակալության ծախսի հաշվարկը

Գործունեության արդյունավետության հիմնական պայմաններից մեկը տարածքի ճիշտ ընտրությունն է: Վարձակալվող տարածքը գտնվում է Երևան ք. բիզնես կենտրոններից մեկում: Այս տարածքում անհրաժեշտ տարածքի վարձակալության ամսական ծախսը կկազմի 300000 դրամ կամ $300000 \cdot 12 = 3600000$ դրամ տարեկան:

7. Ընդհանուր տնտեսական ծախսերի հաշվարկը

Ընդհանուր տնտեսավարման ծախսերի մեջ մտնում են ձեռնարկության ընդհանուր կառավարման-ադմինիստրատիվ՝ գործարանը կառավարող անձնակազմի աշխատավարձի, գործուղման, տպագրական, փոստային-հեռագրային ծախսերը և այլ ծախսեր: Այն որոշվում է հետևյալ բանաձևով՝

$$\text{ԸՏԾ} = U_{\text{հիւ}} \cdot \% \text{ԸՏԾ} / 100,$$

որտեղ՝ %ԸՏԾ - Ընդհանուր տնտեսավարման ծախսերի տոկոսն է %:

$$\text{ԸՏԾ} = 7998480 \cdot 129 / 100 = 10318039.2 \text{ դրամ:}$$

8. Փաթեթի ընդհանուր ինքնարժեքի կալկուլացիան

Ինքնարժեքի կալկուլացիան բերված է աղ. 6.5-ում:

Ինքնարժեքի կալկուլացիան:

N	Ծախսերի հոդվածի անվանումը	գումարը, դրամ
1.	Համալրող առարկաներ	2045000
2.	Էլեկտրաէներգիա	600000
3.	Ընդհանուր հիմնական աշխատավարձ	7998480
4.	Լրացուցիչ աշխատավարձ	952200
5.	Սարքավորումների պահպանման և շահագործման ծախսեր	10277899.68
6.	Տարածքի վարձակալություն	3600000
7.	Ընդհանուր տնտեսական ծախսեր	10318039.2
	Ինքնարժեք	35791618.88

Համակարգի ինքնարժեքը կկազմի՝

 $F = 35791618.88$ դրամ:**9. Շահույթի և միավորի գնի հաշվարկը**

Համակարգի գինն իր մեջ ընդգրկում է ընկերության շահույթն ու ինքնարժեքը:

Շահույթի հաշվարկն իրականացվում է հետևյալ բանաձևով՝

$$\mathcal{C} = F * \mathcal{C}/100,$$

որտեղ F –ն համակարգի ինքնարժեքն է, \mathcal{C} –ն՝ շահույթի դրույքաչափը: Մեր օրինակի համար.

$$\mathcal{C} = 35791618.88 * 18/100 = 6442491.398 \text{ դրամ} :$$

Գնի հաշվարկը իրականացվում է հետևյալ կերպ՝

$$G = F + \mathcal{C}:$$

Մեր օրինակի համար.

$Q = 35791618.88 + 6442491.398 = 42234110.28$ դրամ:

Համակարգի բացթողնման գինը որոշվում է հետևյալ բանաձևով՝

$$Q_{\text{բաց}} = Q + \text{ԱԱՀ},$$

որտեղ $Q_{\text{բաց}}$ –ը համակարգի/փաթեթի բացթողնման գինն է, ԱԱՀ–ն՝ ավելացված արժեքի հարկը (20%):

$$Q_{\text{բաց}} = 42234110.28 + 42234110.28 * 20/100 = 50680932.33 \text{ դրամ:}$$

Այսպիսով, կատարված հաշվարկների արդյունքում ստացանք, որ Էլեկտրոնային փաստաթղթերի (PDF) բովանդակության ճանաչման և ամփոփման ավտոմատացված տեղեկատվական համակարգի ինքնարժեքը կկազմի 35791618.88 դրամ, իսկ գինը՝ 50680932.33 դրամ:

ԵԶՐԱԿԱՑՈՒԹՅՈՒՆՆԵՐ

Հետազոտության հիմնական արդյունքների ամփոփում

Հետազոտության արդյունքում մշակվել է PDF փաստաթղթերի բովանդակության ճանաչման և ամփոփման ավտոմատացված համակարգ, որն օգտագործում է արհեստական բանականության տեխնոլոգիաներ: Համակարգը կարող է մշակել ինչպես տեքստային, այնպես էլ սկանավորված PDF փաստաթղթեր, դուրս բերել դրանցից տեքստը և պատկերները, կիրառել OCR տեխնոլոգիա սկանավորված նյութերի համար, իրականացնել բովանդակության ամփոփում և դասակարգում:

Համակարգը կառուցված է մոդուլային սկզբունքով և ներառում է հետևյալ հիմնական բաղադրիչները՝

- PDF մշակման մոդուլ (pdfplumber և PyMuPDF գրադարաններով)
- Պատկերների ճանաչման մոդուլ (OCR՝ pytesseract գրադարանով)
- Տեքստի ամփոփման և դասակարգման մոդուլ (Google Gemma 3 մոդելի օգտագործմամբ)

Հետազոտության արդյունքում ստեղծվել է ոչ միայն վեբ հիմքով օգտագործողի ինտերֆեյս, այլ նաև API ինտերֆեյս, որը թույլ է տալիս համակարգի ֆունկցիոնալությունը ինտեգրել այլ համակարգերում: Համակարգն աջակցում է բազմալեզու միջավայր՝ հնարավորություն տալով աշխատել հայերեն, ռուսերեն և անգլերեն լեզուներով, ինչպես ինտերֆեյսի, այնպես էլ փաստաթղթի բովանդակության մակարդակում:

Համակարգի առավելությունները և սահմանափակումները

Առավելություններ

1. **Բազմալեզու աջակցություն** - Համակարգը կարող է մշակել և ամփոփել փաստաթղթեր հայերեն, ռուսերեն և անգլերեն լեզուներով՝ պահպանելով բնօրինակ լեզվի առանձնահատկությունները:
2. **Ունիվերսալ մշակում** - Համակարգը կարող է մշակել ինչպես տեքստային, այնպես էլ սկանավորված PDF փաստաթղթեր՝ կիրառելով OCR տեխնոլոգիա:
3. **Մոդուլային կառուցվածք** - Համակարգի մոդուլային կառուցվածքը հեշտացնում է դրա պահպանումը և հետագա զարգացումը:

4. **Ինտեգրման հնարավորություններ** - API ինտերֆեյսը թույլ է տալիս համակարգի ֆունկցիոնալությունը հեշտությամբ ինտեգրել այլ համակարգերում և հավելվածներում:
5. **Բարելավված օգտագործողի փորձառություն** - Համակարգն ունի ինտուիտիվ և հարմարավետ օգտագործողի ինտերֆեյս բազմաթիվ UX բարելավումներով:
6. **Բովանդակության դասակարգում** - Համակարգը առաջարկում է փաստաթղթի բովանդակության ավտոմատ դասակարգում ըստ կատեգորիաների, ինչը հեշտացնում է փաստաթղթերի կատալոգավորումը:
7. **Արդյունքների կառավարման գործիքներ** - Համակարգն առաջարկում է ամփոփման արդյունքների պատճենման և ներբեռնման հնարավորություններ:

Սահմանափակումներ

1. **Լեզվական սահմանափակումներ** - Համակարգը ներկայումս աջակցում է միայն երեք լեզու (հայերեն, ռուսերեն, անգլերեն), ինչը կարող է սահմանափակող գործոն լինել միջազգային մակարդակում կիրառման համար:
2. **OCR ճշգրտություն** - Սկանավորված փաստաթղթերի տեքստի ճանաչման ճշգրտությունը կախված է պատկերի որակից և տեքստի հստակությունից:
3. **Ֆայլի չափի սահմանափակում** - API-ն ունի ֆայլի չափի սահմանափակում (առավելագույնը 16ՄԲ), ինչը կարող է խնդիրներ առաջացնել մեծածավալ փաստաթղթերի մշակման ժամանակ:
4. **Ուսուցման տվյալների կախվածություն** - Ամփոփման և դասակարգման որակը կախված է Google Gemma 3 մոդելի ուսուցման տվյալներից:
5. **Բարդ կառուցվածքներ** - Համակարգը կարող է դժվարանալ բարդ տեսողական կառուցվածքներով (աղյուսակներ, գրաֆիկներ, և այլն) փաստաթղթերի մշակման ժամանակ:

Հետագա զարգացման հեռանկարները

1. **Լեզուների ընդլայնում** - Համակարգի ընդլայնում լրացուցիչ լեզուների աջակցմամբ, ներառյալ եվրոպական և ասիական լեզուները:
2. **OCR համակարգի բարելավում** - OCR տեխնոլոգիայի հետագա կատարելագործում՝ բարելավելով տեքստի ճանաչման ճշգրտությունը տարբեր որակի պատկերներում:

3. **Առաջադեմ AI մոդելների ինտեգրում** - Նորագույն լեզվական մոդելների ներդրում ամփոփման և դասակարգման որակի բարելավման համար:
4. **Կառուցվածքային վերլուծության բարելավում** - Աղյուսակների, գրաֆիկների և այլ կառուցվածքային տարրերի ճանաչման և մշակման հնարավորությունների զարգացում:
5. **Փաստաթղթերի համեմատության ֆունկցիա** - Մեկից ավելի փաստաթղթերի համեմատության և համադրման հնարավորության ավելացում:
6. **Արդյունքների վիզուալիզացիա** - Ամփոփման արդյունքների վիզուալիզացիայի միջոցների մշակում՝ տվյալների ավելի լավ ընկալման համար:
7. **Մոբայլ հարթակների աջակցություն** - Մոբայլ հարթակների համար հատուկ տարբերակների մշակում՝ բջջային սարքերից փաստաթղթերի մշակման համար:
8. **Ավելի բարդ դասակարգման համակարգ** - Դասակարգման համակարգի ընդլայնում՝ ներառելով ենթակատեգորիաներ և հիերարխիկ դասակարգում:
9. **Օգտագործողի հետադարձ կապի ինտեգրում** - Համակարգի ուսուցման և հարմարեցման մեխանիզմների մշակում՝ հիմնված օգտագործողի հետադարձ կապի վրա:
10. **Բովանդակության պահպանման և արխիվացման գործիքներ** - Ամփոփված փաստաթղթերի պահպանման և արխիվացման համակարգի մշակում:

Այսպիսով, PDF փաստաթղթերի բովանդակության ճանաչման և ամփոփման ավտոմատացված համակարգը հանդիսանում է արդյունավետ լուծում փաստաթղթերի մշակման և դրանց բովանդակության արագ ըմբռնման համար: Չնայած առկա սահմանափակումներին, համակարգը ունի զարգացման մեծ ներուժ և կարող է նշանակալի արժեք ներկայացնել ինչպես գիտական, այնպես էլ բիզնես համայնքների համար՝ հատկապես հայալեզու միջավայրում:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. Adobe Systems. (2008). Document management — Portable document format — Part 1: PDF 1.7. https://opensource.adobe.com/dc-acrobat-sdk-docs/standards/pdfstandards/pdf/PDF32000_2008.pdf
2. O'Hara, S. (2019). "PDF Document Processing: A Survey." *Handbook of Document Image Processing and Recognition*, Springer, 657-684. https://doi.org/10.1007/978-0-85729-859-1_21
3. Hassan, T., & Baumgartner, R. (2007). "Table Recognition and Understanding from PDF Files." *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 1143-1147. <https://doi.org/10.1109/ICDAR.2007.4377094>
4. Déjean, H., & Meunier, J. (2006). "A System for Converting PDF Documents into Structured XML Format." *Document Analysis Systems VII*, Springer, 129-140. https://doi.org/10.1007/11669487_12
5. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.). O'Reilly Media. <https://www.oreilly.com/library/view/python-for-data/9781491957653/>
6. Sweigart, A. (2019). Automate the Boring Stuff with Python (2nd ed.). No Starch Press. <https://automatetheboringstuff.com/>
7. Smith, R. (2007). "An Overview of the Tesseract OCR Engine." *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2, 629-633. <https://doi.org/10.1109/ICDAR.2007.4376991>
8. Breuel, T.M. (2008). "The OCRopus Open Source OCR System." *Proceedings of Document Recognition and Retrieval XV*, 6815, 68150F. <https://doi.org/10.1117/12.783598>
9. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT 2019*. <https://doi.org/10.18653/v1/N19-1423>
10. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). "BART: Denoising Sequence-to-Sequence Pre-training for

- Natural Language Generation, Translation, and Comprehension." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871-7880.
<https://doi.org/10.18653/v1/2020.acl-main.703>
11. Lin, C. Y. (2004). "ROUGE: A Package for Automatic Evaluation of Summaries." *Proceedings of the Workshop on Text Summarization Branches Out*, 74-81.
<https://aclanthology.org/W04-1013/>
 12. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). "Text Summarization Techniques: A Brief Survey." *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8(10), 397-405.
<https://doi.org/10.14569/IJACSA.2017.081052>
 13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). "Attention is All You Need." *Advances in Neural Information Processing Systems*, 30, 5998-6008.
<https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
 14. Dabre, R., Chu, C., & Kunchukuttan, A. (2020). "A Survey of Multilingual Neural Machine Translation." *ACM Computing Surveys*, 53(5), 1-38.
<https://doi.org/10.1145/3406095>
 15. Yadav, V., & Bethard, S. (2018). "A Survey on Recent Advances in Named Entity Recognition from Deep Learning Models." *Proceedings of the 27th International Conference on Computational Linguistics*, 2145-2158. <https://aclanthology.org/C18-1182/>
 16. Ruder, S., Vulić, I., & Søgaard, A. (2019). "A Survey of Cross-lingual Word Embedding Models." *Journal of Artificial Intelligence Research*, 65, 569-631.
<https://doi.org/10.1613/jair.1.11640>
 17. Fielding, R. T., & Taylor, R. N. (2002). "Principled Design of the Modern Web Architecture." *ACM Transactions on Internet Technology*, 2(2), 115-150.
<https://doi.org/10.1145/514183.514185>
 18. Masse, M. (2011). REST API Design Rulebook. O'Reilly Media.
<https://www.oreilly.com/library/view/rest-api-design/9781449317904/>

19. Richardson, L., & Ruby, S. (2008). RESTful Web Services. O'Reilly Media.
<https://www.oreilly.com/library/view/restful-web-services/9780596529260/>
20. Nielsen, J. (2020). 10 Usability Heuristics for User Interface Design. Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
21. Boehm, B., & Basili, V. R. (2001). "Software Defect Reduction Top 10 List." *Computer*, 34(1), 135-137. <https://doi.org/10.1109/2.962984>
22. Juristo, N., & Moreno, A. M. (2013). Basics of Software Engineering Experimentation. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4757-3304-4>
23. ISO/IEC 25010:2011. (2011). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. <https://www.iso.org/standard/35733.html>