

AXBOROTLARNI IZLASH TIZIMLARIDA RELEVANSNI BAHOLASH
USULLARI VA ULARNING SAMARADORLIGI

Fayziyeva Mahbuba Rahimjovna

O'zbekiston Milliy pedagogika universiteti va

O'zbekiston xalqaro islomshunoslik akademiyasi professori.

Dottoyev Sayfulla Xamidullayevich

O'zbekiston Milliy pedagogika universiteti dotsenti.

Annotatsiya: *Ushbu maqolada axborotlarni izlash tizimlarida (AIT) relevansni baholashning to'rtta asosiy ko'rsatkichi — Precision, Recall, nDCG va MAP — nazariy va amaliy jihatdan tahlil qilinadi. Shuningdek, matematik formulalar, qiyosiy jadval va amaliy misollar yordamida har bir ko'rsatkichning afzalligi, kamchiligi va qo'llanilish sohasi ochib beriladi.*

Kalit so'zlar: *axborotlarni izlash tizimi; relevans; Precision; Recall; nDCG; MAP; Cranfield paradigmasi; TREC.*

KIRISH

Raqamli axborot hajmining jadal o'sishi zamonaviy axborot jamiyatining asosiy xususiyatlaridan biriga aylandi. Internet tarmoqlari, ilmiy bazalar va elektron kutubxonalarda to'plangan ma'lumotlar miqdori har yili ikki baravar ortib bormoqda [10]. Bunday sharoitda kerakli axborotni tez va aniq topish uchun samarali axborotlarni izlash tizimlari (AIT) talab etiladi.

Axborot izlash tizimlari ning sifatini ob'ektiv baholash zarurati axborot izlash tizimlari nazariyasida alohida yo'nalish — relevansni baholash metodologiyasini — shakllantirdi [2]. 1957–1967 yillarda o'tkazilgan Cranfield eksperimentlari [3] hozirgi kunda ham qo'llaniladigan baholash paradigmasining asosini yaratdi: har bir tizim uchun qat'iy belgilangan test to'plami (corpus), so'rovlar majmuasi va ekspert relevantlik baholashlari qo'llaniladi.

Zamonaviy axborot izlash tizimlari larni baholashda Precision, Recall, nDCG va MAP ko'rsatkichlari eng keng tarqalgan bo'lib, ularni tanlash tizim maqsadiga bog'liq [4]. Ammo ushbu ko'rsatkichlarni qiyosiy tarzda keng o'zbek tilidagi ilmiy adabiyotlarda tahlil qilinishi hali yetarli emas.

Maqolaning maqsadi — Precision, Recall, nDCG va MAP ko'rsatkichlarini IMRAD tuzilishi asosida nazariy-amaliy tahlil qilish va ularni qiyosiy baholash.

Maqolaning vazifalari:

- axborot izlash tizimlari va relevans tushunchalarining nazariy asoslarini yoritish;
- har bir ko'rsatkichning matematik apparatini va hisoblash algoritmini taqdim etish;
- amaliy misollar va qiyosiy jadval orqali ko'rsatkichlar farqini ko'rsatish;
- turli sohalarda qo'llanilishiga oid tavsiyalar ishlab chiqish.

MATERIALLAR VA METODLAR**Relevans tushunchasi va uning turlari**

Relevans (inglizcha relevance — moslik) — hujjatning foydalanuvchi axborot ehtiyojiga qanchalik mos kelishini ifodalovchi markaziy tushuncha [5]. Saracevic [6] relevansning besh turini ajratadi: (1) tizimli — kalit soʻzlar mosligi; (2) mavzuiy — mazmun mosligi; (3) kognitiv — foydalanuvchi bilim darajasiga mosligi; (4) situatsion — aniq vazifaga mosligi; (5) affektiv — emotsional mosligi.

Amaliyotda relevansni baholashning ikki usuli qoʻllaniladi: binar baholash (relevant = 1, relevant emas = 0) va darajali baholash (0–3 yoki 0–4 shkala). Darajali baholash nDCG kabi zamonaviy koʻrsatkichlar uchun asos boʻladi [2].

Cranfield paradigmasi va standart test toʻplamlari

Axborot izlash tizimlari ni baholashning standart usuli Cranfield paradigmasi [3] deb ataladi. Uning uchta asosiy komponenti mavjud:

- Hujjatlar korpusi — qidiruv olib boriladigan maʼlumotlar bazasi;
- Test soʻrovlari — real foydalanuvchi ehtiyojlarini aks ettiruvchi soʻrovlar toʻplami;
- Relevantlik baholashlari (qrels) — ekspert tomonidan belgilangan har bir soʻrov–hujjat jufti uchun relevanslik darajasi.

Keng tarqalgan test toʻplamlari: TREC [7] (NIST, 1992-yildan beri), MS MARCO [8] (1 mln real foydalanuvchi savoli), CLEF (koʻptilli baholash) va NTCIR (Osiyo davlatlari).

Axborotlarni izlash tizimlari arxitekturas

Zamonaviy axborot izlash tizimlari besh asosiy moduldan iborat [1]:

- Hujjatlar korpusi — veb-sahifalar, ilmiy maqolalar, kitoblar;
- Indeksash moduli — tokenizatsiya, stemming, stop-soʻzlarni olib tashlash;
- Soʻrovlarni qayta ishlash — foydalanuvchi soʻrovini tizim formatiga keltirish;
- Relevansni hisoblash — BM25, TF-IDF, BERT kabi algoritmlarga asoslanadi [9];
- Natijalarni tartiblash — topilgan hujjatlarni relevanslik boʻyicha saralash.

NATIJALAR**Precision va Recall koʻrsatkichlari**

Manning va boshqalar [1] taʼrifiga koʻra, qidiruv natijalarini toʻrt kategoriyaga boʻlish mumkin: TP (toʻgʻri musbat), FP (notoʻgʻri musbat), TN (toʻgʻri manfiy), FN (notoʻgʻri manfiy). Ushbu kategoriyalar asosida asosiy koʻrsatkichlar hisoblanadi:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

Precision foydalanuvchiga koʻrsatilgan natijalar orasida foydalilarning ulushini oʻlchaydi [1]. Recall esa tizimning barcha relevant hujjatlarni topish qobiliyatini ifodalaydi. Ushbu ikki koʻrsatkich oʻrtasida teskari mutanosiblik (trade-off) mavjud: birini oshirish koʻpincha ikkinchisini pasaytiradi [4]. F1-score ularning harmonik oʻrtachasi sifatida muvozanatni taʼminlaydi (formula 3).

Precision@K va Recall@K variantlari faqat natijalar ro'yxatining dastlabki K ta elementini hisobga oladi. Veb-qidiruv tizimlarida amalda Precision@10 eng ko'p qo'llaniladi, chunki foydalanuvchilar birinchi sahifadagi natijalarni ko'radi [4].

nDCG (Normalized Discounted Cumulative Gain)

Järvelin va Kekäläinen [2] tomonidan taklif etilgan nDCG ko'rsatkichi darajali relevans va natijalar tartibini bir vaqtda hisobga oladi. Hisoblash uch bosqichda amalga oshiriladi:

$$CG_p = \sum_{i=1}^p rel_i \quad (4)$$

$$DCG_p = \sum_{i=1}^p (rel_i / \log_2(i + 1)) \quad (5)$$

$$nDCG_p = DCG_p / IDC G_p \quad (nDCG \in [0, 1]) \quad (6)$$

Bu yerda rel_i — i-pozitsiyadagi hujjatning relevanslik balli; IDC G — ideal tartibdagi DCG ning maksimal qiymati. $\log_2(i + 1)$ diskont omili yuqori pozitsiyadagi hujjatlarga ko'proq og'irlik beradi [2]. $nDCG = 1$ ideal tartibni, $nDCG = 0$ to'liq mos kelmaslikni bildiradi.

Hisoblash namunasi (0–3 shkala bo'yicha 5 ta natija):

Tizim natijasi — relevanslik ballari: 3, 2, 3, 0, 1

$$DCG_5 = 3/1 + 2/1,585 + 3/2 + 0 + 1/2,585 \approx 6,149$$

Ideal tartib: 3, 3, 2, 1, 0 → $IDC G_5 \approx 6,324$

$$nDCG_5 = 6,149 / 6,324 \approx 0,972$$

MAP (Mean Average Precision)

MAP Precision-Recall egri chizig'i ostidagi yuzani yaqinlashtiruvchi va ko'p so'rovlar bo'yicha umumlashtirilgan ko'rsatkich sifatida keng qo'llaniladi [7]. Hisoblash ikki bosqichda amalga oshiriladi:

$$AP(q) = (1/R) \times \sum_{k=1}^n [Precision(k) \times rel(k)] \quad (7)$$

$$MAP = (1/Q) \times \sum_{q=1}^Q AP(q) \quad (8)$$

Bu yerda R — to'plamdagi barcha relevant hujjatlar soni; Q — so'rovlar soni; $rel(k)$ — k-pozitsiyadagi hujjat relevantligi (0 yoki 1). AP faqat relevant hujjatlar uchraydigan pozitsiyadagi Precision qiymatlarini hisobga oladi [8]. MAP TREC kampaniyalarida qidiruv tizimlarini taqqoslashning standart ko'rsatkichi sifatida qabul qilingan [7].

Hisoblash namunasi (5 ta relevant hujjat, 10 ta natija):

Natijalar: R N R N N R N N N R (R — relevant)

Relevant pozitsiyalarda Precision: $P@1=1,0$; $P@3=0,667$; $P@6=0,5$; $P@10=0,4$

$$AP = (1,0 + 0,667 + 0,5 + 0,4) / 4 = 0,642$$

Ko'rsatkichlarning qiyosiy tahlili

1-jadval [1,2,4] asosida tuzilgan va to'rtta ko'rsatkichni beshta mezon bo'yicha qiyoslaydi.

1-jadval. Relevansni baholash ko'rsatkichlarining qiyosiy tahlili

Ko'rsatkich	Relevans turi	Tartiblash	Murakkablik	Qo'llanish sohasi
Precision	Binar	Yo'q	$O(n)$ – oson	Veb-qidiruv tizimlari
Recall	Binar	Yo'q	$O(n)$ – oson	Tibbiy, huquqiy tizimlar
F1-score	Binar	Yo'q	$O(n)$ – oson	Muvozanatlangan baholash
nDCG	Darajali	Ha	$O(n \log n)$	Tavsiya tizimlari
MAP	Binar	Ha	$O(n)$ o'rtacha	Ilmiy tadqiqotlar, TREC

Amaliy misol [4]: to'plamda 5 ta relevant hujjat. Tizim A: R R N R N N N N R; Tizim B: N R R N R R N N N N. Precision@10 bo'yicha ikkala tizim $4/10 = 0,40$ — teng. Ammo MAP bo'yicha Tizim A ($AP = 0,63$) Tizim B dan ($AP = 0,49$) ustun, chunki Tizim A da relevant hujjatlar ro'yxat boshida. Bu holat Precision ning tartiblashni hisobga olmasligi tufayli yuzaga keladi.

MUHOKAMA

Ko'rsatkichlarni tanlash mezonlari

Qidiruv tizimining maqsadi ko'rsatkich tanlovini belgilaydi. Tibbiy va huquqiy tizimlarda birorta ham relevant hujjat o'tkazib yuborilmasligi kerak, shuning uchun Recall ustunlik qiladi [5]. Veb-qidiruv tizimlarida esa foydalanuvchi odatda birinchi sahifani ko'radi va Precision@10 hamda nDCG@10 ko'proq ma'lumot beradi [4].

nDCG ning asosiy ustunligi — darajali relevansni qo'llab-quvvatlashidir. Bunday baholash hujjatlar orasidagi sifat farqini aniqroq aks ettiradi [2]. MAP esa ko'p so'rovli baholashlarda statik barqarorligi bilan ajralib turadi va TREC [7] kabi standart kompaniyalarda qidiruv tizimlarini solishtirish uchun qo'llaniladi.

Cheklovlar va kelajak yo'nalishlari

Barcha klassik ko'rsatkichlar bir qator cheklovlarga ega: Precision/Recall tartiblashni hisobga olmaydi [1]; MAP binar relevans bilan cheklangan [8]; nDCG IDCG ni hisoblash uchun barcha relevant hujjatlar ro'yxatini talab qiladi [2]. Bundan tashqari, barcha klassik ko'rsatkichlar statik relevans baholashlariga tayanadi, holbuki haqiqiy foydalanuvchi ehtiyojlari ko'pincha dinamik va kontekstga bog'liq bo'ladi [6].

BERT, GPT kabi katta til modellariga asoslangan zamonaviy axborot izlash tizimlari lar [9] semantik relevansni aniqroq aniqlaydi, biroq ularni baholash uchun ham yuqorida tahlil qilingan ko'rsatkichlar qo'llanilmoqda. Crowdsourcing orqali yig'ilgan katta hajmli

relevanslik baholashlari [8] kelajakda ko'rsatkichlarning ishonchliligini oshirishga xizmat qiladi.

XULOSA

Ushbu maqolada axborot izlash tizimlarida relevansni baholashning to'rtta asosiy ko'rsatkichi IMRAD tuzilishi asosida tahlil qilindi. Quyidagi xulosalar chiqarildi:

Precision va Recall tizimning aniqlik va to'liqligini o'lchaydi. F1-score ularni muvozanatlashtirishda qulay vosita hisoblanadi [1].

nDCG — darajali relevans va tartiblashni birgalikda hisobga oladigan yagona ko'rsatkich — zamonaviy tavsiya tizimlari va veb-qidiruv uchun eng mos metriks [2].

MAP — ko'p so'rovli baholashlarda standart mezon bo'lib, TREC kabi xalqaro kompaniyalarda tizimlarni solishtirish imkonini beradi [7].

Amaliyotda bir nechta ko'rsatkichlarni birgalikda qo'llash tavsiya etiladi: Recall (tibbiy/huquqiy), Precision@K va nDCG@K (vab-qidiruv), MAP (ilmiy tadqiqot).

Ushbu tahlil natijalaridan axborot izlash tizimlari larni loyihalash, sinash va takomillashtirish jarayonlarida amaliy qo'llanma sifatida foydalanish mumkin. Kelajak tadqiqotlar kontekstga asoslangan dinamik relevans baholash usullarini ishlab chiqishga qaratilishi maqsadga muvofiq.

FOYDALANILGAN ADABIYOTLAR

[1] Manning C. D., Raghavan P., Schütze H. Introduction to Information Retrieval. // Cambridge University Press. — 2008. — 544 b. ISBN 978-0-521-86571-5.

[2] Järvelin K., Kekäläinen J. Cumulated gain-based evaluation of IR techniques. // ACM Transactions on Information Systems. — 2002. — Vol. 20, No. 4. B. 422–446. DOI: 10.1145/582415.582418.

[3] Cleverdon C. W. The Cranfield tests on index language devices. // Aslib Proceedings. — 1967. — Vol. 19, No. 6. B. 173–194.

[4] Croft W. B., Metzler D., Strohman T. Search Engines: Information Retrieval in Practice. // Pearson Education. — 2015. — 552 b. ISBN 978-0-13-686675-8.

[5] Baeza-Yates R., Ribeiro-Neto B. Modern Information Retrieval: The Concepts and Technology behind Search. 2nd ed. // Addison-Wesley. — 2011. — 913 b. ISBN 978-0-321-41691-9.

[6] Saracevic T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III. // Journal of the American Society for Information Science and Technology. — 2007. — Vol. 58, No. 13. B. 1915–1933. DOI: 10.1002/asi.20682.

[7] Voorhees E. M., Harman D. K. (Eds.) TREC: Experiment and Evaluation in Information Retrieval. // MIT Press. — 2005. — 440 b. ISBN 978-0-262-22073-9.

[8] Campos R. et al. MS MARCO: A Human Generated MACHine Reading COMprehension Dataset. // arXiv preprint. — 2018. — arXiv:1611.09268v3. URL: <https://arxiv.org/abs/1611.09268>.

[9] Robertson S. E., Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond. // Foundations and Trends in Information Retrieval. — 2009. — Vol. 3, No. 4. B. 333–389. DOI: 10.1561/15000000019.

[10] Dauletov, A.; Muminov, B.; Matyakubova, N.; Abdurahmonova, U.; Bakhriyeva, K.; Fayzieva, M. Development of a Software Model for Classification and Automatic Cataloging of Archive Documents. Information **2026**, 17, 341. <https://doi.org/10.3390/info17040341>

[11] Sayfulla, Dottoyev. "Information and methodological support-as a means of intensifying the educational process. Euro-Asia Conferences, 159-161." 2021,

Internet manbalari:

[12] TREC rasmiy sayti. — URL: <https://trec.nist.gov> (murojaat sanasi: 05.04.2025).

[13] Introduction to Information Retrieval — onlayn kitob. — URL: <https://nlp.stanford.edu/IR-book> (murojaat sanasi: 05.04.2025).