

ILMIY MAQOLALARDA HUJJATLARNI ORGANISH METODI XUSUSIYATLAR

Jumanazarova Zuhra

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti tarjima nazariyasi fakulteti ijtimoiy – gumanitar fanlar kafedrası sotsiologiya yo'nalishi talabasi.

<https://doi.org/10.5281/zenodo.17960979>

Annotatsiya. Hujjatning maqsadi ma'lumotni muallifdan o'quvchilarga uzatishni osonlashtirishdir. Hujjatni unda mavjud bo'lgan ma'lumotlarni aniq va samarali talqin qilish uchun loyihalash muallifning vazifasidir. Buning uchun muallif bir qator stilistik vositalardan foydalanishi mumkin. Ushbu maqolada biz hujjatlarning funktsionalligi tushunchasini taqdim etamiz, bu ma'lumotni uzatish jarayonida hujjatlar va ularning tarkibiy qismlarining rolini tavsiflashga harakat qiladi. Hujjatning funktsional tavsifi hujjat turi, uning mo'ljallangan maqsadlari va hujjatlarni avtomatik talqin qilish va olish strategiyalari haqida tushuncha beradi

Kalit so'zlar: Metodologik kalit so'zlar, hujjatlarni o'rganish, hujjatli tadqiqot, tadqiqot metodi, sifat tadqiqoti, arxiv tadqiqoti, kontent tahlili ma'lumotlarni yig'ish, ma'lumotlarni kodlash, rasmiy hujjatlar, shaxsiy hujjatlar, arxiv materiallari, adabiyotlarni o'rganish.

Kirish: Sohadagi eng so'nggi yangiliklarni aks ettirish uchun qayta ko'rib chiqilgan va kengaytirilgan "Hujjatlarning ilmiy ekspertizasi, to'rtinchi nashr" - bu sud-tibbiy hujjatlarni ekspertiza qilishning eng yaxshi amaliyotlarini batafsil bayon qiluvchi qulay va oson jild.

1989-yilda nashr etilgan birinchi nashrdan beri sud-tibbiy hujjatlarni ekspertiza qilish sohasida keskin o'zgarishlar yuz berdi - ham professional ekspert uchun mavjud bo'lgan analitik usullardan foydalanishdan tortib, ofis va bosmaxona uskunalari va siyohlardagi texnologiyalardagi o'zgarishlargacha. So'roq qilingan hujjatni ishlab chiqarishda ishlatiladigan har qanday materialni, masalan, siyoh yoki qog'oz parchasini tahlil qilishdan maqsad, uni so'roq qilingan hujjatning boshqa joyidagi yoki boshqa hujjatdagi boshqa material bilan taqqoslash, ularning umumiy kelib chiqishi bor-yo'qligini aniqlashdir. Shuningdek, tergovchiga hujjatning mumkin bo'lgan kelib chiqishi haqida ma'lumot berish zarurati tug'ilishi mumkin.

Ushbu so'nggi nashr so'nggi 10-15 yil ichida sodir bo'lgan ko'plab o'zgarishlar va yutuqlarni aks ettiradi. Qamrab olingan mavzular quyidagilarni o'z ichiga oladi: qo'lyozma talqini bo'yicha hozirgi fikrlar; qo'lyozmani tasodifiy va qasddan o'zgartirish; namunalarni to'g'ri to'plash; maydalangan hujjatlarni muhokama qilish; professional akkreditatsiya standartlari, malaka va o'qitish; va dasturiy ta'minot va tasvirlash yordamida hujjatlar va qo'lyozmalarni zamonaviy raqamli tasvirlash va tahlil qilish, jumladan, o'chirish, yo'q qilish va boshqa hujjatlarni o'zgartirish usullaridan tasvirni qayta tiklash. Yangi bo'lim kognitiv tarafkashlikka bag'ishlangan va 8-bob mavjud texnologiyalarga asoslangan holda bosma va nusxa ko'chirilgan hujjatlardagi yutuqlar va bunday hujjatlarni taqqoslashdagi analitik ishlanmalarni qamrab olish uchun to'liq yangilangan.

Hujjatlar to'plamiga asoslangan ilmiy xaritalashda hujjat-hujjat o'xshashligini turli yo'llar bilan o'lchash mumkin. Bir yondashuv hujjatlar bilan bog'liq matnlar o'rtasidagi o'xshashlikni o'lchashdir. Ikkala matnda ham uchraydigan atamalarga asoslangan bir nechta o'xshashlik o'lchovlari Boyce va boshqalar, 1994, Salton va McGill, 1983 adabiyotlarida taklif qilingan. Fan xaritalashda ushbu matnga asoslangan yondashuvdan foydalanadigan ishlarning so'nggi misollari Glenisson, Glänzel, Janssens, and Moor (2005) va Janssens, Leta, Glänzel va

Moor (2006)¹ bo'lib, ularda bibliometriya va kutubxona va axborot fanlari mos ravishda xaritalangan.

Ikkinchi yondashuv bibliografik bog'lanish bo'lib, u 1960-yillarda Kessler, 1963a, Kessler, 1963b, Kessler, 1965 da bir qator maqolalar orqali ilmiy jamiyatga kiritilgan². Bibliografik bog'lanish yondashuvining boshlang'ich nuqtasi ikkita hujjatning ma'lumotnomalar ro'yxatining o'xshashligini o'lchashdir. Ikki hujjat o'rtasida umumiy bo'lgan bitta havola ular orasidagi bog'lanish birligi deb hisoblangan va ikkita hujjat o'rtasidagi bog'lanish kuchi bog'lanish birliklari soni sifatida aniqlangan (Kessler, 1963a). Shubhasiz, ikkita hujjat o'rtasidagi bog'lanish birligi ikkalasi tomonidan ham keltirilgan va bibliografik bog'lanish hujjat-hujjat o'xshashligiga iqtibosga asoslangan yondashuvdir. Bu yerda biz ikkita hujjatning bog'lanish kuchi 0 dan katta bo'lsa, bibliografik jihatdan bog'langanligini aytamiz. Bog'lanish kuchi hujjat-hujjat o'xshashligining ko'rsatkichi sifatida olinadi.

Matnga asoslangan yondashuv bibliografik bog'lanish bilan birlashtirilishi yoki integratsiya qilinishi mumkin. Shunday qilib, bu turdagi yondashuvlar ikkita axborot manbasidan, matn mazmunidan va iqtibos qilingan havolalardan foydalanadi. Integratsiyaga turli usullar bilan, masalan, bir xil hujjatlar juftligi bilan bog'liq ikkita farq qiymatlarining statistik kombinatsiyasi orqali erishish mumkin³.

O'xshashlik qiymatlari olinganda, asosiy yondashuvdan qat'i nazar, hujjatlar klasterlash texnikasini qo'llash orqali avtomatik ravishda juftlik bo'yicha disjunkt to'plamlarga guruhlanishi mumkin. Ushbu ishda, avvalgi maqolada (Ahlgren va Jarneving, 2008) keltirilgan tadqiqot ishlab chiqilgan bo'lib, biz ilmiy xaritalash kontekstida hujjat-hujjat o'xshashligiga beshta yondashuvni eksperimental ravishda taqqoslaymiz. Biz matnga asoslangan yondashuvlarni, iqtibosga asoslangan bibliografik bog'lanish yondashuvini va matnga asoslangan yondashuvlar va bibliografik bog'lanishni birlashtirgan yondashuvlarni taqqoslaymiz. To'liq bog'lanish usuli qo'llanilganda va ikki turdagi o'xshashliklar, birinchi va ikkinchi darajali bo'yicha, yondashuvlar sinov hujjatlarimizning asosiy haqiqat sub'ekti tasnifiga qanchalik mos kelishini tekshiramiz⁴.

Metod va materiallar: Hujjat mantig'ini tahlil qilish geometrik maket tahlili kabi katta e'tiborga sazovor bo'lmadi, ammo sahifalarni tushunishning ba'zi qiziqarli usullari taklif qilingan.

Ularning eng katta qismi qoidalarga asoslangan texnikalar bo'lib, ularda ikkita katta usul sinfini aniqlash mumkin: Daraxt o'zgarishlariga asoslangan usullar va Formatlash bilimlariga asoslangan usullar. Birinchi sinfga tegishli texnikalar geometrik maket daraxtini daraxt ichidagi tugunlarni harakatlantirishga va har bir tugunni ma'lum qoidalar to'plamiga muvofiq tegishli mantiqiy yorliq bilan izohlashga harakat qiladi. Bunday usulning misoli ko'p maqolali tushunishdir. Hujjat xususiyatlari bilan bog'liq bo'lgan mantiq qoidalarini optimallashtirish uchun sahifaning joylashuvi haqidagi dastlabki bilimlarga asoslangan boshqa yondashuv qo'llaniladi.

¹ Xanno Valishevski: "Fazoviy hujjatlarni talqin qilish uchun avtomatik bilimlarni olish", 4-stajyor materiallarida. Hujjatlarni tahlil qilish va tanib olish bo'yicha konferensiya, 1997-yil avgust, 243-247-betlar.

² Xuhong Li, Peter A. Ng: "O'rganish qobiliyatiga ega hujjatlarni tasniflash va ajratib olish tizimi", 5-xalqaro Hujjatlarni tahlil qilish va tanib olish konferensiyasi materiallarida

³ A. Antonakopulos va A. Koenen: "Tesseral tasvir yordamida mintaqa tavsifi va qiyosiy tahlil", 5-xalqaro Hujjatlarni tahlil qilish va tanib olish konferensiyasi materiallarida

⁴ M. Vishvanatan: "Skanerlangan hujjatni tahlil qilish - Sintaktik yondashuv",

Formatlash bilimlari usullariga misollar sifatida “sahifadagi eng mos bloklarni tavsiflash yoki sintaktik grammatika qo‘llanilishi” keltiriladi. Yaqinda chop etilgan maqola bizning vazifamizga qaratilgan, ammo turli usullardan foydalanadi.

Bizning tizimimiz nafaqat bir nechta yondashuvlarning yangi kompleks kombinatsiyasi, balki yangi strategiyalar va g‘oyalarni birlashtiradi. Avvalo, biz nafaqat dagi kabi geometrik ma’lumotlardan foydalanamiz, balki qo‘shimcha ravishda biz matnli ma’lumotlardan (masalan, matn naqshlari), shrift ma’lumotlaridan (masalan, uslub, o‘lcham, qalinlik va boshqalar) va OCR tomonidan berilgan universal atributlardan (masalan, qatorlar soni) foydalanamiz. Boshqa yondashuvlarning eng muhim kengaytmasi - bu turli joylashuv obyektlari o‘rtasida aniqlanishi mumkin bo‘lgan munosabatlarni ifodalash qobiliyatidir (masalan, muallif (to‘g‘ridan-to‘g‘ri) sarlavha ostiga joylashtiriladi)⁵.

Fizik nuqtai nazardan, hujjat bir nechta sahifalardan iborat. Sahifa bir nechta bloklardan, blok bir nechta satrlardan va satr bir nechta so‘zlardan iborat. Hujjatda har bir turdagi kamida bitta obyekt bo‘lishi kerak.

Mantiqiy nuqtai nazardan, hujjatda mantiqiy birliklar, masalan, sarlavha, jadvallar yoki ro‘yxatlar va boshqalar mavjud. Mantiqiy birlik bir nechta xarakterli ko‘rinishga ega: u sarlavha, sarlavha, referat, muallif, izoh, sarlavha va boshqalar bo‘lishi mumkin. Bunday hollarda mantiqiy birliklar bitta blokdan iborat. Bu mantiqiy birliklargahar biri ehtimollik qiymatiga ega bo‘lgan turli yorliqlar berilishi mumkin (masalan, 80% sarlavha, 12% sarlavha, 8% muallif).

Lekin mantiqiy birliklar, shuningdek, ro‘yxat yoki jadvalga bloklarning agregatlari bo‘lishi mumkin. Bu ikki turdagi har biri bitta yorliqqa ega bo‘ladi (mos ravishda 100% ro‘yxat yoki jadval). Ro‘yxat (bir o‘lchovli) bloklar ketma-ketligidan iborat bo‘lsa-da, ularning o‘zlari ro‘yxat elementlari (yoki kichik ro‘yxatlar) deb belgilangan bo‘lsa-da, jadvallar yanada murakkabroq: ular hatto ba’zi nosimmetrikliklar bilan birga (ikki o‘lchovli) massivga o‘xshash tuzilmadir.

Muhokama: Ushbu tadqiqotda biz ilmiy xaritalash kontekstida hujjat-hujjat o‘xshashligiga beshta yondashuvni ko‘rib chiqdik. Yondashuvlardan ikkitasi matnga asoslangan, bitta yondashuv, bibliografik bog‘lanish, iqtibosga asoslangan, qolgan ikkita yondashuv esa matnga asoslangan yondashuvlar va bibliografik bog‘lanishni birlashtiradi. Biz to‘qqizta usulni eksperimental ravishda taqqosladik, bu yerda har bir usul ilmiy xaritalash kontekstida beshta yondashuvdan biri bilan aniqlanadi yoki ulardan biriga kiradi.

Hujjat tuzilishini aniqlash ikki xil ishlov berish bosqichiga bo‘linadi: birinchidan, sarlavhalar, izohlar va ro‘yxatlar kabi barcha umumiy hujjat tuzilmalari tan olinadi va ikkinchidan, barcha domenga bog‘liq mantiqiy elementlar tegishli qoidalar bazasi yordamida tan olinadi⁶.

Birinchi bosqich uchun ro‘yxat ichidagi sanab o‘tilgan ro‘yxat ob’ektlarini, jadval katakchalarini jadvallar ichida ifodalash yoki satrlarni sarlavhalar yoki altbilgilarga umumlashtirish uchun ko‘proq aniqlik talab qilinadi.

Ikkinchi bosqich uchun qoidalar bazasining barcha qoidalari geometrik, shriftga xos va matnli xususiyatlardan foydalangan holda barcha bloklarga nisbatan mustaqil ravishda tekshiriladi. Agar qoida ishga tushsa, blok tegishli yorliqni oladi.

Bloklar na qayta segmentlanmaydi, na ba’zi pastki obyektlar qayta tartiblanmaydi.

⁵ <https://www.solix.com/uz/products/answers/ai-for-document-analysis/>

⁶ Y. Wang, T. Philips va R. Haralick: "Tasvirdan SGML/XML tasvirlashgacha: bitta usul", Hujjatlarni joylashtirishni talqin qilish va uning qo‘llanilishi bo‘yicha stajyor materiallar

Mantiqiy yorliqlash uchun hujjat ichidagi har bir (jismoniy) blok alohida tekshiriladi va har bir blok uchun qoida tarjimoni testlari qo'llaniladi, haqiqiy qoida bazasidagi qoidalardan biri ishlayaptimi yoki yo'qmi. Bu holda, blokga tegishli mantiqiy yorliq beriladi. Ushbu yondashuv bilan blok faqat bittadan ko'proq yorliq olishi mumkin, masalan, "80% sarlavha, 20% sarlavha".

Bu qancha va qaysi qoidalar ushbu blokga ishlaganiga bog'liq⁷.

Muqobil yorliqlarni baholash uchun biz ehtimollik yondashuvidan foydalanamiz. Har bir yorliqning reytinglari belgilangan qoida atributlarining og'irligiga (6-bobga qarang) va har bir atributning moslik ishonchligiga bog'liq. Atributlar qanchalik ko'p va "yaxshiroq" bajarilsa, yorliqning reytingi shuncha yuqori bo'ladi.

Natijada, har bir atribut mosligi loyqa moslikdir. Masalan, blok nafaqat "qalin" yoki "qalin emas" yoki u faqat "belgilangan mintaqada" emas, balki so'zlarning aksariyati qalin va ba'zilar muntazam bo'lsa, u holda "qalin bo'lishi kerak" sharti bilan qoida taxminan 90% ga mos keladi.

Yoki agar blok aniq belgilangan mintaqada bo'lmasa, "xy mintaqasida bo'lishi kerak" sharti ma'lum bir tezlik bilan mos keladi. Barcha belgilangan atributlarning har bir loyqa qiymatlari yakuniy reytingga erishish uchun o'rtacha funksiya bilan umumlashtiriladi⁸.

Xulosa va natijalar: Ushbu bo'limda biz tajriba natijalarini taqdim etamiz. Sinovdan o'tkazilgan to'qqizta hujjat-hujjat o'xshashlik usullarining har biri uchun va berilgan o'xshashlik turi (birinchi darajali yoki ikkinchi darajali) bo'yicha mos keladigan klaster yechimi va asosiy haqiqat tasnifi o'rtasidagi kelishuv o'lchandi. 1-jadvalda kelishuv qiymatlari (Rand indeksi), shuningdek, har bir klaster yechimi uchun klasterlar soni ko'rsatilgan. Masalan, jadvalning pastki qismidan uchinchi qatorni ko'rib chiqing. Ushbu maqolada biz bir nechta sohalariga moslashtiriladigan almashtiriladigan qoidalar bazasiga ega hujjat tuzilishini aniqlash uchun qoidaga asoslangan yondashuvni taqdim etdik. Qoidalar matnli va geometrik joylashuv xususiyatlarini o'z ichiga oladi. Ular nafaqat o'z-o'zidan bog'liq atributlardan foydalanadilar, balki bir nechta bloklar orasidagi munosabatlarni ifodalash uchun o'zaro bog'liq atributlardan ham foydalanadilar. Bundan tashqari, qoida birliklari har qanday mantiqiy ifoda bilan murakkab shartli qismga ega bo'lgan "if-then" qoidasiga birlashtiriladi. Bu xususiyatlar sinovdan o'tgan sohalarimizdagi barcha mantiqiy obyektlarni ifodalash uchun yetarli kuch beradi⁹.

Qoidalarning murakkabligi kuchli xususiyatdir, ammo ayni paytda uning eng katta kamchiligidir. Hozirgacha qoidalarni yaratish, sinovdan o'tkazish va moslashtirish inson tomonidan yaratilgan va juda ko'p vaqt talab qiladigan vazifadir. Buning uchun vizualizatsiya vositasi foydalanuvchiga sichqonchani bosish orqali ijobiy yorliq misoli orqali qoidani avtomatik ravishda aniqlashga yordam beradi. Ushbu qoida yorliqlash prototipining ish vaqtida boshqa yorliq misollariga osongina o'zgartirilishi va moslashtirilishi ("o'rganilishi") mumkin. Keyingi qadamimiz qoidalarni asosiy haqiqat hujjatlari to'plamidan o'rganish mumkin bo'lgan tarzda qoidalarni o'rganish qobiliyatini amalga oshirish bo'ladi.

⁷ S. J. Simske: "XML va XML-ma'lumotlardan foydalanib, hujjatlarni jismoniy, mantiqiy va taqdimot darajalarida tushunish", Hujjatlarni joylashtirishni talqin qilish va uning qo'llanilishi

⁸ M. Junker, A. Abecker: Hujjatlarni tasniflash uchun murakkab o'rganish naqshlari", AAAI/ICML Hujjatlarni tasniflash bo'yicha o'rganish

⁹ K. Summers: "Mantiqiy hujjat tuzilishini avtomatik ravishda kashf etish", doktorlik dissertatsiyasi, Kornell universiteti, 1998-yil avgust

Ushbu vazifa uchun bir nechta strategiyalar tayyorlanadi va taqqoslash va/yoki ovoz berish tegishli hujjat sohasi uchun eng yaxshisini ko'rsatadi¹⁰.

Loyqa moslashtirish natijalari ba'zi qoidalar o'rtasidagi aralashuvdan xulosalar chiqarish imkonini beradi. Mantiqiy obyektlar guruhlarining aralashuvi o'xshashliklarni aniq ko'rsatadi, ular sifat jihatidan ifodalanishi mumkin, masalan, "abstrakt-body" "matn-body" ga qanchalik o'xshashligini ko'rish uchun. Buning uchun maxsus mezonlar va o'xshashlik o'lchovlarini ishlab chiqish kerak.

WWWda mos hujjat tasviri va qidiruvi tobora muhimlashib borayotgani va barcha kerakli ma'lumotlar hujjatlarni tahlil qilish va tushunish bosqichlarida to'planganligi sababli, (HTML)/XML formatida hujjat chiqishini taqdim etish g'oyasi o'z-o'zidan ravshan. Shunday qilib, skanerlangan hujjat tasviridan Internetdagi mos taqdimotga to'g'ridan-to'g'ri o'zgartirish mumkin bo'ladi. Ushbu vazifaning asosiy maqsadi hujjat tasvirida ishlatiladigan barcha strukturaviy va barcha joylashuv xususiyatlarini (shrift o'lchamlari, masofalar va boshqalar) o'z ichiga olgan aniq 1:1 taqdimotni taqdim etish bo'lishi kerak¹¹.

Taklif va tafsiyalar: Har bir yorliq uchun bitta qoida aniqlanadi. Yorliqlash vazifasining murakkabligi tufayli qoida tuzilishi oddiy. Qoidalar faqat domenga bog'liq obyektlarni topish uchun ishlatiladi. Umumiy hujjat tuzilmalarini (sarlavhalar, ro'yxatlar va boshqalar) aniqlash oldingi bosqichda qattiq kodlangan algoritmlar yordamida amalga oshiriladi. Birinchi qism aslida tekshirilgan blok va boshqa blok o'rtasidagi geometrik munosabatlarni ifodalaydi, masalan, "Boshqa blok (ehtimol, ma'lum atributlar yoki yorliqlar bilan) malakali mintaqada joylashgan bo'lishi kerak"¹².

Bu mintaqalar kuzatilgan obyektga nisbatan yoki hujjat koordinatalari orqali malakalanishi mumkin.

Nisbiy pozitsiyalar Allen vaqt munosabatlarining ikki o'lchovli kengaytmasi yordamida aniqlanadi [1][3].

Shunday qilib, "A blok B ostida bo'lishi kerak" yoki "A B ning o'ng tomonida bo'lishi kerak, lekin B dan past bo'lmasligi kerak" kabi pozitsiyalarni ifodalash mumkin.

Keng qamrovli tajribalarimiz davomida biz Allen munosabatlari bilan ushbu sifat spetsifikatsiyalari masofalarni aniqlash qobiliyatiga ega emasligini angladik. Shunday qilib, masalan, "sarlavha stol ostida joylashgan bo'lishi kerak" deb ifodalash mumkin emas, chunki bu "ostida" masofasidan qat'i nazar, har qanday obyektga qo'llanilishi mumkin.

Shunday qilib, biz sifat jihatini masofa o'lchovlari diapazonlari bilan kengaytirdik.

Yuqoridagi misolni quyidagicha kengaytirish mumkin: "Sarlavha stol ostida joylashgan bo'lishi kerak; masofa 1 dan ortiq bo'lishi va sarlavha satri balandligining 4 barobaridan oshmasligi kerak".

Ikkinchi qism ikkala blok ichidagi umumiy so'zlarning mutlaq yoki nisbiy sonini yoki tegishli blokda joylashgan bo'lishi kerak bo'lgan oldindan belgilangan so'zlar ro'yxatini belgilaydi. Bu qoidalarni ta'riflashga imkon beradi, masalan: "Qoida qo'llanilishi mumkin bo'lgan blok bilan kamida bitta umumiy so'zga ega bo'lgan blok bo'lishi kerak va bundan tashqari, u "Hurmatli" va "Janob" so'zlarini o'z ichiga olishi kerak". Bu, masalan, xatning oluvchi maydoni va salomlashish maydoni o'rtasidagi munosabatni ifodalashi mumkin.

¹⁰ https://api.scienceweb.uz/storage/publication_files/6911/18955/65ab9be9a513e___hujjatlarni%20tahlil%20qilish%20maqola.pdf

¹¹ Golub: "txt2html - Matnni HTMLga o'zgartiruvchi", <http://www.aigeeek.com/txt2html/>

¹² <https://www.emagia.com/uz/resources/glossary/automated-document-parsing/>

Nihoyat, o'zaro bog'liq blokni ma'lum bir yorliq orqali aniqlash mumkin. Yorliq sharti juda foydali, ayniqsa, bir-biriga bog'liq bo'lgan hujjat obyektlariga nisbatan. Bu atribut juda qiziqarli effektini keltirib chiqaradi:

Bitta mantiqiy (yorliqlangan) obyektning paydo bo'lishi boshqa obyektlarning yorlig'lanishini keltirib chiqaradi! Masalan, agar "sarlavha" ilmiy maqolada topilsa, unda "muallif" (yoki "abstract") odatda to'g'ridan-to'g'ri "sarlavha" ostida topiladi. Shunday qilib, agar (noma'lum) hujjatda "sarlavha" mavjud bo'lmasa, hech qanday "muallif" (yoki "abstract") ham mavjud bo'lmaydi va obyektning noto'g'ri, masalan, "muallif" deb belgilash orqali aniqlikning pasayishi xavfi kamayadi¹³.

Ammo bu xususiyat ham muammoga olib kelishi mumkin: agar "muallif" yorlig'i "sarlavha" blokiga bog'liq bo'lsa, unda "muallif" (yoki "abstract") "sarlavha" topilmaguncha hech qachon yorliqlanmaydi. Agar "sarlavha" obyekti o'tkazib yuborilsa, unda boshqa barcha bog'liq yorliqlar tan olinmaydi.

Foydalanilgan adabiyotlar ro'yxati:

1. Xanno Valishevski: "Fazoviy hujjatlarni talqin qilish uchun avtomatik bilimlarni olish", 4-stajyor materiallarida. Hujjatlarni tahlil qilish va tanib olish bo'yicha konferensiya, 1997-yil avgust, 243-247-betlar.
2. ¹ Xuhong Li, Peter A. Ng: "O'rganish qobiliyatiga ega hujjatlarni tasniflash va ajratib olish tizimi", 5-xalqaro Hujjatlarni tahlil qilish va tanib olish konferensiyasi materiallarida
3. A. Antonakopulos va A. Koenen: "Tesseral tasvir yordamida mintaqa tavsifi va qiyosiy tahlil", 5-xalqaro Hujjatlarni tahlil qilish va tanib olish konferensiyasi materiallarida
4. M. Vishvanatan: "Skanerlangan hujjatni tahlil qilish - Sintaktik yondashuv",
5. <https://www.solix.com/uz/products/answers/ai-for-document-analysis/>
6. Y. Wang, T. Philips va R. Haralick: "Tasvirdan SGML/XML tasvirlashgacha: bitta usul", Hujjatlarni joylashtirishni talqin qilish va uning qo'llanilishi bo'yicha stajyor materiallar
7. S. J. Simske: "XML va XML-ma'lumotlardan foydalanib, hujjatlarni jismoniy, mantiqiy va taqdimot darajalarida tushunish", Hujjatlarni joylashtirishni talqin qilish va uning qo'llanilishi
8. ¹ M. Junker, A. Abecker: Hujjatlarni tasniflash uchun murakkab o'rganish naqshlari", AAAI/ICML Hujjatlarni tasniflash bo'yicha o'rganish
9. K. Summers: "Mantiqiy hujjat tuzilishini avtomatik ravishda kashf etish", doktorlik dissertatsiyasi, Kornell universiteti, 1998-yil avgust
10. https://api.scienceweb.uz/storage/publication_files/6911/18955/65ab9be9a513e_hujjatlarini%20tahlil%20qilish%20maqola.pdf
11. Golub: "txt2html - Matnni HTMLga o'zgartiruvchi", <http://www.aigeeek.com/txt2html/>
12. <https://www.emagia.com/uz/resources/glossary/automated-document-parsing/>
13. R. Brugger: "Hujjatlarni tuzishning statistik usuli", doktorlik dissertatsiyasi,
14. S. Tsujimoto, H. Asada: "Ko'p maqolali hujjatlarni tushunish", 10-chi Naqshni aniqlash bo'yicha xalqaro konferensiya materiallarida, Atlantik-Siti, Nyu-Jersi, 1990, 551 - 556-betlar

¹³ R. Brugger: "Hujjatlarni tuzishning statistik usuli", doktorlik dissertatsiyasi,