



yaratishi mumkin bo'lgan foydalanuvchilar sonini hisobga olgan holda, bunday ma'lumotlarni aniqlash va tasniflashning an'anaviy usullaridan foydalanish qiyin ko'rinadi. Ta'riflangan tadqiqotning maqsadi - IAD modellari, usullari va algoritmlari asosida dasturiy ta'minot prototipini ishlab chiqish orqali veb-sahifalarni tasniflashning aniqligini oshirish. Dasturiy ta'minot prototipi quyidagi talablarga javob berishi kerak: o'qitilgan model veb-sahifani bashorat qilish aniqligi 70% dan ortiq bo'lgan 14 toifadan biriga tasniflashi kerak.

Asosiy qism: Web-sahifalar tasnifining oddiy matndan asosiy farqi gipermatndir. Shu munosabat bilan tasniflashni maqsadli veb-sahifadagi tarkibga qarab tasniflash mumkin. Web-sahifalarning atributlarini matnli va vizualga bo'lish mumkin [3]. Matn ma'lumotlari tasniflashda foydalanish uchun qulayroqdir. Buning uchun bir nechta atributlarni tanlash variantlari qo'llaniladi, masalan, so'zlar sumkasi, TF-IDF va n-gram. Bunday usullar odatda matn qazib olish tadqiqotlarida qo'llaniladi. Veb-sahifa matnni o'z ichiga olishi mumkin bo'lgan konteyner sifatida HTML teglaridan foydalanadi. Bunday teglarni atributlar sifatida tanlash mumkin [4]. Har bir teg uchun "vaznlar" [5], [6] dan foydalanib, tasniflash aniqligini oshirish mumkin. Veb-sahifalar navigatsiya, kontent va boshqa bloklar kabi vizual elementlar ierarxiyasi bilan ifodalanishi mumkin [7]. Bunday usulni amalga oshirish har doim ham aniqlikni oshirmaydi. Natijada, tasniflashning aniqligini oshirish uchun ushbu yondashuvlar birlashtirilishi mumkin [8]. Bundan tashqari, ma'lumotlarning o'lchovlilikini kamaytirish orqali klassifikator [3] ish faoliyatini yaxshilash mumkin. Har bir veb-sahifaning o'ziga xos URL manzili mavjud bo'lib, undan butun veb-sahifani yuklab olmasdan nisbatan tez tasniflash [11] amalga oshirilishi mumkin. Natija an'anaviy matn tasnifiga qaraganda ancha past bo'lsa-da. Veb-sahifaning URL tasnifida [12] n-gramdan foydalanish samaradorlikni oshirishi mumkin. Ta'lim tasnifi modellari uchun ma'lumotlarni tayyorlash. IAD yordamida ushbu maqsadga erishish uchun ma'lumotlarni to'plash va filtrlash va mashinani o'rganish (ML) usuli yordamida tasniflash modelini o'rgatish kerak. IDA usullaridan foydalangan holda muammolarni hal qilishning eng muhim bosqichlaridan biri bu o'quv va test namunalarni yig'ishdir. Model o'zgartirilishi sababli, ma'lumotlarni ma'lum parametrlarning ta'sirini o'rganish uchun mavjud bo'lishi uchun saqlash kerak. Bunday holda, faylda saqlash qulay emas, chunki munosabatlardan foydalanish kerak bo'ladi va ma'lumotlar bazalarida (MB) bu juda sodda tarzda amalga oshiriladi. Yig'ishda "xom" ma'lumotlarni saqlash tavsiya etiladi, bu veb-sahifalarning manba kodi; Dastlab, g'oya nafaqat maqsadli veb-sahifalarni, balki veb-sahifa ichidagi havolalar orqali kirish mumkin bo'lgan ularning pastki sahifalarini ham to'plash edi. Bu juda ko'p vaqt talab qiladigan ish ekanligi ma'lum bo'ldi. Ma'lumot to'plashni boshlashdan oldin, ma'lumotlarni olishingiz mumkin bo'lgan manbaga kirishga arziydi. Ko'pincha "e'tirozli" yoki taqiqlangan tarkibni topishning yagona yo'li - bunday faoliyat bilan bevosita shug'ullanish.

Veb-sahifalardagi matnli ma'lumotlardan tasniflash modellarini o'rgatish.



Olingan ma'lumotlarni qayta ishlash uchun ko'plab mashhur vositalar mavjud: RapidMiner, MATLAB, Python. Har birining imkoniyatlari taxminan bir xil, shuning uchun tanlov foydalanuvchiga bog'liq. Biz Python va ochiq scikit-learn (sklearn) ML kutubxonasidan foydalanishga qaror qildik. Sklearn kutubxonasida matn bilan ishlashda uni raqamlashtirish va vektor sifatida taqdim etish odatiy holdir, ular uchun 3 ta sinf mavjud: HashingVectorizer, CountVectorizer, TfidfVectorizer. Vektorlashtirishning asosiy parametrlarini atributlar soni (n_features yoki max_features) va n_gram deb atash mumkin. Dastlab biz HashVectorizer dan atribut chegarasi 20 000 va n_gram (1, 2) dan foydalanamiz. Trening uchun biz Random Forest algoritmidan foydalanamiz. Tasniflashning aniqligini o'lchash uchun biz sklearn-da allaqachon amalga oshirilgan o'zaro tekshirishdan foydalanamiz. Hisoblash uchun chalkashlik matritsasi ishlatiladi. tp rost musbat, tf rost yolg'on bo'lsin, u holda ikkilik tasniflashning aniq formulasi quyidagicha ko'rinadi:

$$P = \frac{tp}{tp+tf} \quad (1)$$

Ko'p sinfli tasniflashda makro-o'rtacha va mikro-o'rtacha atamaları qo'llaniladi [14]:

$$P_{macro} = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i+fp_i}}{l}, \quad (2)$$

$$P_{micro} = \frac{\sum_{i=1}^l tp_i}{\sum_{i=1}^l (tp_i+fp_i)} \quad (3)$$

qaerda l- sinflar. Macro-averaging har bir sinfga olingan ko'rsatkichda teng vazn beradi, micro-averaging esa har bir hujjatga teng og'irlik beradi. Agar sinflarning og'irligi bir xil bo'lsa, xatoning narxi nuqtai nazaridan, macro-averaging, aks holda - micro-averaging va test to'plamiga ushbu sinfning ko'proq hujjatlarini qo'shish mantiqiy. Bunday holda, sinflar xatolik qiymati bo'yicha teng emas, shuning uchun biz baholash uchun aniq micro-averaging ko'rsatkichdan foydalanamiz. Harflardan tashqari barcha belgilar matnini shablon yordamida tozalaymiz: "[^a-zA-Z]", o'qitilgan modelning aniqligi 0,660. Filtrlashning yana bir usuli - stemming. Agar biz NLTK kutubxonasidan foydalansak, biz 0,665 aniqlikdagi modelni olamiz. Biz to'xtash so'zlari bo'yicha filtrlash qobiliyatini amalga oshiramiz - precision_micro 0.690. Biz barcha filtrlardan foydalanamiz, so'ngra "ing|ly|ed|ious|ies|ive|es|s|ment" so'zlarining oxirini olib tashlaymiz. Bunday holda, biz 0,697 aniqlikni olamiz. Keling, tasniflash algoritmini almashtirishga va modelni sinab ko'rishga harakat qilaylik. SVM, Logistic Regression, kNN, Decision Tree algoritmlarini ko'rib chiqaylik. Natijalar: - kNN - 0,432; - SVM - 0,725; - Logistik regressiya - 0,758; - Qarorlar daraxti - 0,645. Parametrlarni tanlashni odatda empirik deb atash mumkin. Klassifikatorning aniqligi vektorlashtirishga ham, tasniflash algoritmiga ham bog'liq. Keling, meta-klassifikatorlarni qo'shish orqali modelni yaxshilashga harakat qilaylik. Eng yaxshi natijani Random Forest bilan Bagging algoritmi ko'rsatdi: 0,729. Ko'p toifali tasniflash muammosini ko'rib chiqayotganimiz sababli, biz "birga qarshi dam olish" strategiyasidan foydalanamiz. Ushbu strategiya sklearn kutubxonasida ham amalga



oshiriladi. Natijada, modelning aniqligi: 0,763. Ushbu bosqichda Random Forest algoritmidan Bagging meta-algoritmi va one versus rest strategiyasi bilan birgalikda foydalanish eng yaxshi natijaga erishish imkonini berishi aniqlandi. Olingan modelning aniqligi amaliy foydalanish uchun etarli emas, shuning uchun biz veb-sahifalarni tasniflash va o'zimizni taklif qilish uchun mavjud usullardan foydalangan holda uni yaxshilashga harakat qilamiz. Keyinchalik, barcha algoritmlar har bir bosqichda qayta qo'llanildi. Shu bilan birga, Logistic Regression algoritmi yuqori aniqlik darajasiga ega edi, shuning uchun undan foydalanish natijalari asos sifatida olinadi. O'quv ma'lumotlarini tayyorlashning eng muhim bosqichlaridan biri vektorlashtirishdir. Bu holda asosiy parametrlar atributlarni olish algoritmi, so'zlar soni (n-gram) va atributlarning maksimal soni. Sklearn kutubxonasi ma'lumotlarni vektorlashtirishning bir nechta ilovalarini o'z ichiga oladi: HashingVectorizer (HV) - kiritilgan matnni matndagi so'zning takrorlanish soni uchun qiymatlar bilan vektorga aylantirish; TfidfVectorizer (TF) - so'zning takrorlanish sonining hujjatdagi so'zlarning umumiy soniga nisbati qiymatlari bilan vektorga aylantirish. TF-IDF-da qo'shimcha parametr sifatida siz standart TF sonini almashtiradigan va hisob-kitoblardan "oddiy" so'zlarni olib tashlashga imkon beruvchi sublinear funktsiyadan (SB - sublinear) foydalanishingiz mumkin. Natijalar jadvalda ko'rsatilgan.

Atributlar soni	HV	TF	TF+SB
n-gram(1,2)			
5000	747	789	790
20 000	767	778	781
max	767	772	778
n-gram(1,3)			
5000	774	788	790
20000	757	783	780
50000	758	772	779
n-gram(2,3)			
5000	529	687	699

(1)

Eng yuqori aniqlik 5000 atributli TF-IDF+sublinear va n-gramm (1, 2) bilan vektorlashtirish yordamida erishiladi. Internetdan ma'lumotlarni yuklab olishda asosiy muammo shundaki, ma'lumotlar juda ko'p va barcha veb-sahifalar to'g'ri belgilanganmi yoki yo'qmi, domen ijarasi muddati tugaganmi yoki yo'qligini aniqlash juda qiyin. Shu sababli, veb-sahifalarni "to'g'ri tasniflangan" bo'yicha filtrlashga qaror qilindi. Buning uchun biz butun namunani vektorlashtiramiz, so'ngra olingan vektor bo'yicha mashq qilamiz va undan foydalanib bashorat qilamiz. Bunday holda, noto'g'ri prognoz qilingan veb-sahifalar soni 5% dan kam bo'lishi mumkin. Shuning uchun biz ularni kirish imkoni bo'lmagan veb-saytlar yoki dastlab xatoga yo'l qo'ygan toifalar uchun olamiz va ularni o'quv namunasiidan chiqarib tashlaymiz. Aniqlik 0,858 ga oshdi. [3]-moddada PCA (asosiy komponent usuli) dan foydalanish imkoniyati ko'rsatilgan. Bu tahlil qilingan ma'lumotlar to'plamini hal qilinayotgan muammo nuqtai nazaridan optimal bo'lgan hajmga kamaytirish imkonini beradi. Ushbu usul ma'lumotlarni tasniflash uchun tayyorlash uchun ishlatiladi. Turli xil algoritmlar va parametrlar bilan



eng yaxshi natija 0,875 ni tashkil qiladi. Olingan natijalar aniqlikning oshishini ko'rsatadi, ammo kelajakda usul tadqiqot natijalariga ta'sir qilmaydi.

Web-sahifalarni teglar asosida tasniflash.

IADda eng muhim omil - bu atributlardan to'g'ri foydalanish. Veb-sahifalarda siz [2]-[4] teglarni atributlar sifatida tanlashingiz mumkin, masalan, "sarlavha", "a", "meta" va selektorlar qabul qilingan "sarlavha", "tarkib" va "kolontitr" bloklari. teg, id va sinf bo'yicha. Ushbu matn veb-sayt ishlab chiquvchilari tomonidan ta'kidlanganligi sababli, ehtimol ular asosiylari. Har bir atribut uchun alohida tasniflagichlardan foydalanish aniqlikni oshirmaydi. Keling, ularni mavjud so'zlar bilan birlashtirishga harakat qilaylik va shu bilan kalit so'zlarning paydo bo'lish sonini oshiramiz. Natijada, "sarlavha" va "meta" teglarini (tavsif, kalit so'zlar) qo'shish 0,767 aniqlikni berishi eksperimental tarzda hisoblab chiqildi.

URL asosida veb-sahifalarni tasniflash.

Tasniflashning aniqligini oshirish uchun keling, boshqa atributlardan foydalanishga harakat qilaylik. Bir nechta ishlar [9], [10] veb-sahifani tasniflash uchun URL manzilidan (Uniform Resource Locator) foydalanish imkoniyatini ko'rsatdi. URL - bu internetdagi manba manzilini yozishning standartlashtirilgan usuli. Internetda veb-saytlar soddalashtirilgan URL formatidan foydalanadi: <scheme>://<host>:<port>/<path>?<parameters>#<anchor>. Har bir veb-sahifa o'zining noyob URL manzilidan (manzil) foydalanadi va bu ishlab chiquvchilar uchun tushunarli manzillarni - qisqartmalar yoki noaniq identifikatorlarni emas, balki o'qilishi mumkin bo'lgan so'zlarni o'z ichiga olgan manzillarni yaratish uchun "yaxshi shakl" ning norasmiy qoidasidir. Shu munosabat bilan, URL manzilidan foydalanish nafaqat tasniflashning tezkor usulidir. Asosiy kalit so'zlarni veb-sahifa manzilida topish mumkin. Keling, ushbu imkoniyatdan foydalanamiz va manzilni maxsus belgilar va raqamlardan, shuningdek, kengaytmalar, birinchi darajali domenlar va protokollar kabi deyarli barcha manzillarda mavjud so'zlardan filtrlaymiz. Odatda manzil ajratilmasdan bir nechta so'zlardan iborat bo'ladi, shuning uchun biz n-gram belgisidan foydalanamiz. Tegishli parametrlarni tanlagandan so'ng, 0,631 aniqlik olindi. Aniqlik ancha past, bu kutilmoqda, chunki manzillar qisqartmalar, qisqartmalar yoki sozlangan so'zlardan iborat bo'lishi mumkin. Keling, veb-sahifalarning yana bir xususiyatidan foydalanamiz - matnning asosiy g'oyasini matn shaklida saqlaydigan sarlavha. Sarlavha bo'yicha tasniflasangiz, aniqlik 0,669 ga etadi. Keling, atributlarni birlashtiramiz va shu bilan alohida atributlarning etishmayotgan kalit so'zlarini qo'shamiz. Biz 0,713 aniqlikni olamiz.

Ikki darajali kalit so'zlar asosida veb-sahifalarni tasniflash.

Tasniflashning yana bir qiyinligi - bu bir vaqtning o'zida ko'plab toifalarga tegishli bo'lishi mumkin bo'lgan matnning tegishliligini aniqlash, ammo mavjud bo'lganlarning birortasiga emas, masalan, "yangiliklar". Birinchidan, "yangiliklar" toifasini qo'shamiz. Yangiliklar, moliya, sport, siyosat, siyosat, sog'liqni saqlash, texnologiya, texnologiya, madaniyat, san'at, ob-havo, iqtisod, biznes, turmush tarzi, dunyo, milliy, sayohat,



mashhur, filmlar kabi so'zlarni kalit so'zlar sifatida qidiramiz. musiqa, moda. Kalit so'zlardan foydalanish yangi emas va boshqa toifalar qanday aniqlanganidan deyarli farq qilmaydi. Shuning uchun biz baholash va hisoblash mezonini kiritamiz - agar asosiy kalit so'zning kamida ikkita takrorlanishi bo'lsa, biz hisob-kitob qilamiz, bu holda "yangiliklar". Shunday qilib, kalit so'zlarning 2 darajasi mavjud bo'lib, ularda asosiysi birinchi darajada, qolgan so'zlar esa ikkinchi darajada. Moslashish foizini oshirish uchun biz sinonimlardan foydalangan holda to'xtash so'zlarining ko'p o'lchovli vektorini olamiz - bu bizga bir xil ma'noga ega so'zlarni hisoblamasdan so'zlarning paydo bo'lishining to'g'riroq foizini olish imkonini beradi. Noaniqlik holatida xato darajasini pasaytirish. Tasniflashda har bir toifa bo'yicha tasniflash natijalarini baholashga imkon beradigan xatolik narxi matritsasi juda tez-tez ishlatiladi. Ko'p sinfli tasnifda yana bir muammo to'g'ri bo'lmaganda toifani tanlashdir, shuning uchun o'qitish "noma'lum" ga topshiriq yordamida amalga oshirildi. Xatolik darajasini pasaytirishning yana bir usuli - bashorat qilish algoritmiga ovoz balandligini tekshirishni qo'shishdir. Agar o'lcham etarlicha kichik bo'lsa, veb-sahifaning toifasi "noma'lum" bo'ladi. Ushbu veb-sahifalar stublar yoki matnni o'z ichiga olmaydigan veb-sahifalar bo'lishi mumkin. Shuningdek, aniqlikni oshirish uchun biz ma'lum toifalarga ega domenlar ro'yxatidan foydalanishni qo'shamiz. Bunday echimlarni mavjud namunada sinab ko'rish mumkin bo'lgan ortiqcha moslama tufayli qiyin - domenlar ro'yxati o'quv namunasi olingan va hajm bo'yicha tekshirish faqat namunadagi sahifalar alohida toifa sifatida mavjud bo'lganda mumkin.

Ierarxik tasniflash usuli. "Birga qarshi dam olish" strategiyasidan foydalanish model sifatini yaxshilash imkonini berdi. Shuning uchun biz shunga o'xshash usuldan foydalanamiz, farqi shundaki, biz alohida tanlangan parametrlar va algoritmgaga ega modellardan foydalanishimiz mumkin. Buning uchun, keling, ularning toifalari uchun qat'iy o'qitilgan yagona ikkilik tasniflagichlarni olaylik. Yakuniy toifani olish uchun biz ovoz berish usulidan foydalanamiz. Klassifikatorlarning chiqish ma'lumotlarini yaxlitlash orqali yuqori aniqlikka erishish mumkin edi. Treningsiz "to'g'ri bashorat qilingan"larga ko'ra tasniflashda aniqlik 0,879 ni tashkil etdi. Trening bilan "to'g'ri bashorat qilingan" bo'yicha tasniflashda - 0,928.

Xulosa. Maqolada veb-sahifalarni tasniflashning ma'lum usullarining ko'pchiligi ko'rib chiqildi va o'zimiznikilar ham taklif qilindi. 14 toifadan birini aniqlash qobiliyatiga ega bo'lgan va 96% mikro-o'rtacha aniqligi bilan veb-sahifalarni tasniflash modeli ham taklif qilingan. Olingan modelning aniqligi ideal emas. Veb-sahifalarni tasniflashda asosiy qiyinchilik matnni o'z ichiga olmaydi, balki tasvir va videolarni ham o'z ichiga olgan sahifalardir. Odatda, odam kontentni vizual tarzda baholaydi. Shu sababli, ushbu turdagi atributlarni qo'shish tasniflash sifatini yaxshilashga yordam beradi, shuning uchun tasniflash tadqiqotida mumkin bo'lgan yo'nalishlardan biri yangi turdagi atributlardan, masalan, tasvirlardan foydalanish bo'lishi mumkin.

**ADABIYOTLAR RO'YXATI:**

1. Rukavitsin A.N. Izv ma'lumotlarini qazib olish usullaridan foydalangan holda veb-sahifalarni tasniflash modelini ishlab chiqish. SPbSETU "LETI". 2016 yil. 4-son. B.12-20.
2. Sample Overview: alQaida-Like mobile discussions & potential creative uses // Site of Federation Of American Scientists. URL: <http://fas.org/irp/eprint/mobile.pdf>. (Дата обращения 15.11.2015).
3. Qi X., Davison B. D. Web page classification: Features and algorithms // J. ACM Computing Surveys. 2009. Vol. 41, iss. 2, № 12.
4. A fuzzy system for the web page representation / A. Ribeiro, V. Fresno, C. M. Garcia-Alegre, D. Guinea // Intelligent exploration of the Web. Physica-Verlag HD. 2003. P. 19-37.
5. Kwon O. W., Lee J. H. Web page classification based on k-nearest neighbor approach // IRAL '00 Proc. of the fifth Intern. workshop on Information retrieval with Asian languages, Hong Kong, China, 2000. P. 9-15.
6. Kwon O. W., Lee J. H. Text categorization based on k-nearest neighbor approach for Web site classification // Inform. Process. Manage. 2003. Vol. 39, № 1. P. 25-44.
7. Visual Adjacency Multigraphs – a Novel Approach for a Web Page Classification / M. Kovacevic, M. Diligenti, M. Gori, V. Milutinovic // Proc. of SAWM 2004 workshop, ECML 2004, New York, NY USA, 2004.
8. Web-page classification through summarization / D. Shen, Z. Chen, Q. Yang, H. J. Zeng, B. Zhang, Y. Lu, W. Y. Ma // Proc. of the 27th annual Intern. ACM SIGIR conf. on Research and development in information retrieval, UK, Sheffield, 2004. P. 242-249.
9. Qi X., Davison B. D. Web page classification: Features and algorithms // Computing Surveys. 2009. Vol. 41, № 2. P. 12.
10. Belmouhcine A., Benkhalifa M. Implicit Links based Web Page Representation for Web Page Classification // Proc. of the 5th Intern. conf. on Web Intelligence, Mining and Semantics. Larnaca, Cyprus, 2015. P. 12.
11. Kan M. Y., Thi H. O. N. Fast webpage classification using URL features // Proc. of the 14th ACM Intern. conf. on Information and knowledge management. Bremen, Germany, 2005. P. 325-326.
12. Abdallah T. A., Iglesia B. URL-based web page classification-a new method for URL-based web page classification using n-gram language models // SCITEPRESS Digital Library-KDIR 2014-Intern. conf. on Knowledge Discovery and Information Retrieval. Rome, Italy, 2014. P. 14.
13. Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks // Information Processing & Management. 2009. Vol. 45, № 4. P. 427-437.



14. Learning word vectors for sentiment analysis / A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts // Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011. Vol. 1. P. 142–150.



YUZNI ANIQLOVCHI SUN'IY INTELLEKTLAR VA ULARNING ISHLASHI

Suyunov E.

CHDPU 1-kurs magistri

Yuldasheva G.T.

Ilmiy rahbar, CHDPU dotsent v.b.

Annotatsiya: *Maqolada yuzni aniqlovchi sun'iy intellekt, ularning turlari va yuzni aniqlash texnologiyasi qanaqa ishlashi haqida ma'lumot berilgan.*

Kalit so'zlar: *sun'iy intellekt, yuzni aniqlash, biometrik, recognition, yuzni tanib olish, iris.*

Yuzni tanib olish - bu shaxsning yuzidan foydalangan holda shaxsini aniqlash yoki tasdiqlash usuli. Yuzni tanish tizimlari fotosuratlar, videolar yoki real vaqtda odamlarni aniqlash uchun ishlatilishi mumkin.

Yuzni tanib olish biometrik xavfsizlik kategoriyasidir. Biometrik dasturiy ta'minotning boshqa shakllariga ovozni aniqlash, barmoq izini aniqlash va ko'zning to'rt pardasi yoki irisini aniqlash kiradi. Texnologiya asosan xavfsizlik va huquqni muhofaza qilish uchun ishlatiladi, ammo foydalanishning boshqa sohalariga qiziqish ortib bormoqda.

Yuzni tanish maxsus kameralar yonidan o'tayotgan odamlarning yuzlarini kuzatuv ro'yxatidagi odamlarning suratlariga moslashtirish orqali ishlaydi. Kuzatuv ro'yxatlarida har qanday odamning, shu jumladan, hech qanday huquqbuzarlikda gumon qilinmagan odamlarning suratlari bo'lishi mumkin va tasvirlar istalgan joydan, hatto bizning ijtimoiy tarmoqlardagi akkauntlarimizdan ham olinishi mumkin. Yuz texnologiyalari tizimlari har xil bo'lishi mumkin, lekin umuman olganda, ular quyidagicha ishlaydi:

1-qadam: Yuzni aniqlash

Kamera yolg'iz yoki olomon ichida yuz tasvirini aniqlaydi va joylashtiradi. Rasmda to'g'ridan-to'g'ri yoki profilga qaragan odam ko'rinishi mumkin.

2-qadam: Yuz tahlili

Keyinchalik, yuzning tasviri olinadi va tahlil qilinadi. Yuzni tanish texnologiyasining aksariyati 3D tasvirlarga emas, balki 2D tasvirlarga tayanadi, chunki u 2D tasvirni ommaviy suratlar yoki ma'lumotlar bazasidagilar bilan qulayroq moslashtirishi mumkin. Dastur yuzingizning geometriyasini o'qiydi. Asosiy omillar orasida ko'zlaringiz orasidagi masofa, ko'z bo'shlig'ining chuqurligi, peshonadan iyakgacha bo'lgan masofa, yonoq suyaklarining shakli, lablar, quloqlar va iyak konturi kiradi. Maqsad - yuzingizni farqlash uchun asosiy bo'lgan yuz belgilarini aniqlash.

3-qadam: Tasvirni ma'lumotlarga aylantirish

Yuzni suratga olish jarayoni analog ma'lumotni (yuzni) shaxsning yuz xususiyatlariga asoslangan raqamli ma'lumotlar (ma'lumotlar) to'plamiga aylantiradi.