Ekaia, 2021, 39, 327-336
https://doi.org/10.1387/ekaia.21879

ekaia
ZIENTZIA eta TEKNOLOGIA ALDIZKARIA

ISSN 0214-9001 – eISSN 2444-3255

Web-Scraping Teknikan Oinarritutako Azpiegitura Informatikoak. *Xerka Online* eta *Minerva* aplikazioak

(*Web-Scraping* Based Informatics Infrastructures. Applications: *Xerka Online* and *Minerva*)

Borja Fernandez-Gauna, Naiara Rojo, Unai Fernandez-Gamiz*

Vitoria-Gasteizko Ingeniaritza Eskola, Euskal Herriko Unibertsitatea (UPV/EHU)

LABURPENA: Erabiltzailearen zereginak errazten dituzten sistema informatiko ugari erabiltzen dira, bai arlo profesionalean, bai eta arlo pertsonalean ere. Hala ere, kasu batzuetan erabiltzaileen beharren eta sistema informatikoak eskaintzen duenaren arteko distantzia handia da. Artikulu honetan *web-scraping* teknikarekin sortutako bi azpiegitura informatiko deskribatu dira, jatorrizko beste azpiegitura batzuen funtzionalitatea hobetu dutenak. Alde batetik, *Xerka Online* aplikazioak ikertzaileen *curriculum vitae*aren (CVaren) sortze- eta mantentze-lana errazten du, ikertzaileek egin behar izaten duten ataza nagusia modu automatizatuan eginez: argitalpenak bilatu eta haiei dagozkien kalitate-adierazle (eragin-faktore eta aipamen kopuru) eguneratuak ezarri. *Minerva* aplikazioak, ordea, Vitoria-Gasteizko Ingeniaritza Eskolan egiten diren kalitate-txostenak kudeatzen ditu. Horretarako, Euskal Herriko Unibertsitateko (UPV/EHUko) GAUR web-aplikaziotik automatikoki jaisten ditu itxitako aktak, jarritako kalifikazioen estatistikak kalkulatzen ditu, eta maila ezberdinetan egiten diren txostenak batzen ditu. Bi aplikazioen abantaila nagusiak lan horiek egiteko behar den denboraren eta giza akatsen murrizpena dira.

HITZ GAKOAK: azpiegitura, web-scraping, bilatu, argitalpenak, GAUR, UPV/EHU.

ABSTRACT: *Computational systems that facilitate the tasks of the customer are frequently used, both for professional and personal purposes. However, in some cases, the computer system does not meet the users' needs. In this article, two computational infrastructures based on the use of* web-scraping *are described, which improve the functionality of the original infrastructures.* Xerka Online *allows the creation and maintenance of a researcher's* curriculum vitae *(CV) by searching his/her publications and their updated quality indicators (impact factor and cites). Minerva manages the quality assessment reports in the Faculty of Engineering of Vitoria-Gasteiz. To that end, it downloads the grade records from GAUR (a web-application of the University of the Basque Country UPV/EHU), calculates statistics, and merges reports generated in different levels of the quality assessment process. The main advantages of these applications are time reduction and avoidance of human errors.*

KEYWORDS: infrastructure, web-scraping, search, publication, GAUR, UPV/EHU.

**** Harremanetan jartzeko / Corresponding author:*** Borja Fernandez-Gauna. Vitoria-Gasteizko Ingeniaritza Eskola, Euskal Herriko Unibertsitatea (UPV/EHU), Nieves Cano kalea, 12 (01006 Vitoria-Gasteiz). – borja.fernandez@ehu.eus – https://orcid.org/0000-0001-9233-2333.

Nola aipatu / How to cite: Fernandez-Gauna, Borja; Rojo, Naiara; Fernandez-Gamiz, Unai (2021). «*Web-Scraping* Teknikan Oinarritutako Azpiegitura Informatikoak. Xerka Online eta Minerva aplikazioak»; *Ekaia*, 39, 2021, 327-336. (https://doi.org/10.1387/ekaia.21879).

Jasoa: 2020, ekainak 30; Onartua: 2020, irailak 10.

ISSN 0214-9001 - eISSN 2444-3255 /

1. **SARRERA**

Zerbitzu informatikoak eskaintzeko modu ezberdinak daude: besteak beste, mahai gainerako aplikazioak (*Windows*, *Linux* edo *MacOs* sistema eragilea duen konputagailuan erabiltzeko), eskuko telefonorako aplikazioak (*Android* edo *iOS* sistema eragilea duen gailuan erabiltzeko), edo web-aplikazioak (web-nabigatzaile baten bidez erabiltzen direnak). Azken motako aplikazioek abantaila argia dute: edozein sistema eragiletatik edota gailutatik atzi daitezke, baldin eta sistema eragileak web-nabigatzailea badu. Horretaz gain, garatze-kostua, oro har, beste bi aplikazio motarena baino baxuagoa dela esan daiteke, kode komuna erabil daitekeelako erabiltzaile guztientzat. Hori dela eta, web-aplikazioak erabiltzen dira eremu askotan.

Oro har, web-aplikazioek zerbitzari-bezero egitura erabiltzen dute. Alde batetik, zerbitzariak sistemaren informazioa eta aplikazioaren logika kudeatzen ditu. Beste aldetik, erabiltzaileak web-nabigatzailea (bezeroa) erabiltzen du aplikazioaren interfazea bistaratzeko eta aplikazioarekiko elkarrekintza kudeatzeko. Web-aplikazio gehienek *Hypertext Markup Language* (HTML) lengoaia erabiltzen dute interfazea definitzeko, eta *Hypertext Transfer Protocol* (HTTP) protokoloa zerbitzari-bezero komunikazioa egiteko.

HTML lengoaia testuan oinarritzen da, eta dokumentuak definitzeko balio du. Horretan, elementu bakoitzari dagokion meta-informazioa gehitzeko etiketa ezartzen zaio. Adibidez, dokumentu baten titulua definitzeko, *title* etiketa erabiltzen da. Beraz, dokumentu jakin baten titulua *Nire orriaren titulua* bada, hori honela adieraziko litzateke: *<title>Nire orriaren titulua</title>*. HTML-k bi objektu nagusi definitzen ditu erabiltzailearen eta aplikazioaren arteko elkarrekintza ahalbideratzeko: estekak eta formularioak. Estekek erabiltzaileari aukera ezberdinak eskaintzeko balio dute, eta formularioek, berriz, testua, zenbakiak edota beste edozein datu mota eskatzeko.

Bestalde, HTTP eta *Hypertext Transfer Protocol Secure* (HTTPS) protokoloek testu-mezuetan oinarritutako hainbat eragiketa definitzen dituzte zerbitzariaren eta bezeroen artean komunikatzeko. HTTP protokoloak edozeinek irakur ditzakeen testu hutseko mezuak erabiltzen ditu, eta HTTPS protokoloak segurtasun-geruza gehitzen dio mezuari, hori enkriptatuz. Bai HTTP eta bai HTTPS protokoloak hiru zatiko mezuak erabiltzen dituzte: a) eskatutako helbidea edo *Uniform Resource Locator* (URL), b) goiburuak (meta-datuak bidaltzeko, adibidez, jatorrizko helbidea edo web-nabigatzailearen bertsioa), eta c) mezuaren edukia. Protokoloek definitutako eragiketen artean bi nabarmentzen dira, oso erabiliak direlako:

— GET. Bezeroak URLa eskatzen du, eta eskaeraren parametroak URL helbidean bertan kodetzen dira. Adibidez, *https://www.google.es* bi-

latzailean «kaixo» hitza bilatzeko, web nabigatzaileak *https://www.google.es/search?q=kaixo* URL helbidea eskatuko dio zerbitzariari. Zerbitzariak parametroa deskodetuko du eskaera jasotzean eta, behin erantzuna prest dagoenean, HTML dokumentuaren bidez erantzungo dio.

— POST. Bezeroak URLa eskatzen du, eta eskaeraren parametroak mezuaren edukian kodetzen dira. Aurreko adibidean, nabigatzaileak *https://www.google.es* helbidea eskatuko du, eta mezuaren edukia *q=kaixo* izango da.

HTTP/HTTPS protokoloetan datu iraunkorrak erabiltzeko mekanismoa *cookie*ak dira. *Cookie*ak bezeroak gordetzen dituen testu-fitxategiak dira. Zerbitzariak eskaera jakin bati erantzutean, *cookie* bat gordetzeko eska diezaioke bezeroari, eta honek, erabiltzaileak horien erabilera onartzen badu, aurrerantzean bueltan bidaliko dio *cookie* hori zerbitzariari. Mekanismo horren bitartez, zerbitzariak bezeroa identifika dezake, eta informazio iraunkorra gorde.

1.1. *Web-scraping*-a eta haren erabilera

Web-scraping software programek erabiltzen duten teknika da, web-orrietatik informazioa modu automatikoan (erabiltzaileak parte hartu gabe) lortzeko [1, 2, 3]. *Web-scraping* erabiltzen duen aplikazioari *web-scraper* deritzo. Datu-bilketa automatikoa egiteko, bezeroak programatikoki simulatzen du gizakiaren eta zerbitzariaren arteko elkarrekintza, hori web-nabigatzailearen bidez egingo balitz bezala. Hau da, programa batek egiten dio HTTP/HTTPS eskaera zerbitzariari, HTML erantzuna jasotzen du, eta, dokumentua bistaratu gabe, behar duen informazioa erantzunetik ateratzen du, agindutako zeregina bete arte.

Edozein HTML dokumentutatik informazioa erraz atera daiteke, dokumentuaren elementu bakoitza dagokion etiketaren bidez identifikatuta dagoelako [4, 5, 6]. Horrela, erabiliko den dokumentuaren egitura aztertu ostean, *web-scraper*ra kodetu eta horrek edozein elementu dagokion etiketaren bidez bila dezake. Teknikaren arazo nagusia garatutako aplikazioren mantentze-lana da. Web-scraping algoritmoak dokumentu-egitura eta urrats-sekuentzia jakinean oinarritzen dira, eta ondorioz, egitura edo urratsak aldatuz gero, web-scraperra bera eguneratu beharko da. Zorionez, administrazioarekin lotutako web-aplikazioak egonkorrak izaten dira, eta ez dira maiz eguneratzen. Horrek murriztu egiten du nolabait mantentze-lanaren kostua. Lan honetan aurkeztu diren bi kasuetan, ez da oraindik arrazoi horrengatik behin ere software-a eguneratu behar izan.

Web-scrapingaren erabilera nagusia era automatizatuan datuak saretik lortzea da. Helburuak askotarikoak izan daitezke; hauek, esaterako: merka-

tuko prezioak eguneratuta mantentzea [7, 8], merkatuen joera aztertzea [9], erabakiak hartzeko laguntza eskaintzea [10], sare sozialetan gertaera baten aurrean erreakzioak aztertzea, enpresa ezberdinek eskaintzen dituzten prezio eta zerbitzuen alderatzea, eta abar.

Artikulu honetan web-scraping-a dagoeneko existitzen diren sistema informatikoen erabilera hobetzen duten sistema berriak eraikitzeko erabili da. Teknika hainbat azpiegituraren gainean sortutako bi azpiegitura garatzeko erabili da: *Xerka Online* eta *Minerva*. Haien helburua oinarrizko azpiegituren funtzionalitatea osatzea eta hobetzea izan da, erabiltzaileari esperientzia hobea eskaintzeko asmoz.

2. *WEB-SCRAPING*-aren BI APLIKAZIO: *XERKA ONLINE* ETA *MINERVA*

Azken urteotan, tresna, programa eta aplikazio informatikoak etengabe sortzen dira, ahal den neurrian eguneroko lanak automatizatzeko, gure beharrak asetzeko edota zerbitzu berriak emateko. Artikulu honetan irakasle/ikertzaileen esperientzia hobetzeko bi tresnaren garapena deskribatzen da: *Xerka Online* eta *Minerva*. Lehenak argitalpen zientifikoen datu-baseak miatzeko web-aplikazioak erabiltzen ditu CVaren parte garrantzitsua betetzeko. Bigarrenak, berriz, Vitoria-Gasteizko Ingeniaritza Eskolan urtero egiten diren kurtso-amaierako txostenak automatikoki betetzen ditu. Bi aplikazioek erabiltzaileen denboraren aurrezpen nabaria ahalbidetzen dute, eta, gainera, giza akatsak murrizten dituzte.

2.1. *Xerka Online*

Unibertsitateko irakasle eta ikertzaile guztiek, beren lanbide-karreran zehar, hainbat egoera ezberdinetan aurkeztu behar izaten dute beren CVa edo haren atal bat; adibidez, ikerketa-proiektuetan parte hartzeko eskaera ofizialak egitean edota ikerketa-jarduera ebaluatzeko (seiurtekoak) eskabideak betetzean. Nabarmentzekoa da, halaber, irakaskuntza-jarduera ebaluatzeko egiaztapenak eskatzean (ANECAri edo Unibasq-i) CV xehatua aurkeztu behar dutela, eta horrek denbora luzea eskatzen diola irakasle/ikertzaileari. Oro har, lanbide-karreran irakasle/ikertzaile guztiek eskatu behar dute behin baino gehiagotan egiaztapena, bai lanpostu ez-iraunkorra lortzeko (irakasle atxikiaren lanpostua lortzeko, adibidez) bai lanpostu iraunkorra lortzeko ere (irakasle agregatua edo titularra, adibidez). Horretaz aparte, lanpostu iraunkorra izanik ere lan-kategoria hobetu nahi dutenek egiaztapen-eskaera egin beharko dute. Espainiako Hezkuntza eta Lanbide Heziketaren Ministerioaren azken datuen arabera [11], unibertsitate publikoetan eta pribatuetan 125.471 irakasle/ikertzaile kontratatu zeuden 2018/2019 ikastaroan, eta horietatik % 54,01ek besterik ez zuen kon-

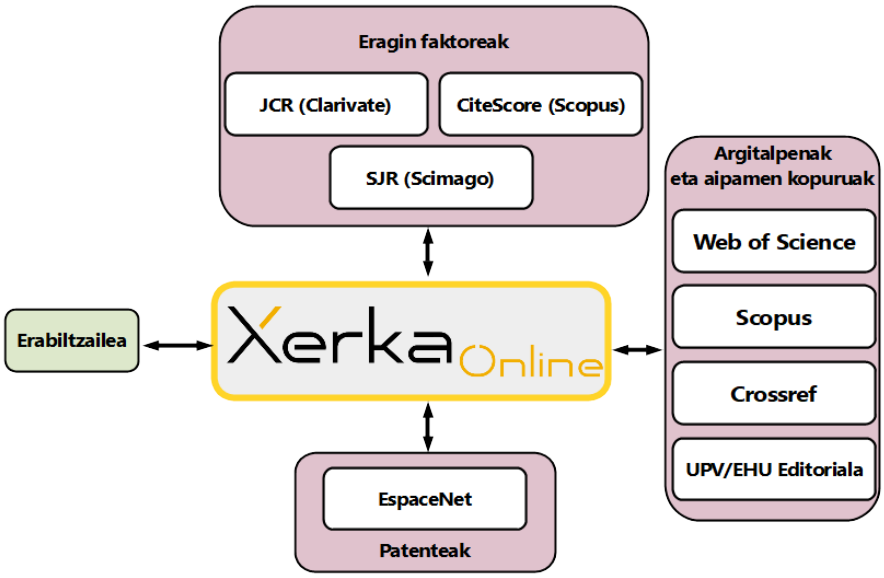
tratu iraunkorra. Euskadin, urte horretan, 5.620 irakasle/ikertzaile zeuden kontratatuta (4.475 Euskal Herriko Unibertsitatean UPV/EHU), horietatik % 56,16 kontratu iraunkorrarekin. Beraz, guztira denbora luzea inbertitzen da prozesu honetarako CVa osatzen eta eguneratzen.

Irakasle/ikerlatzaileak hainbat meritu mota aurkezten ditu bere CVan, hala nola, prestakuntza-, irakaskuntza-, ezagutzaren hedapen-, ikerketa-, berrikuntza-jarduerak. Meritu horietako bakoitzaren kasuan hainbat datu jaso behar dira CVan, eta, gainera, eskatutako CV-eredua deialdi mota batetik bestera aldatu egiten da. Beraz, CVaren eguneratzeak eta moldatzeak lan handia eskatzen dio irakasle/ikerlariari.

Ikerketa-jarduerari dagokion CVaren ebaluazioa egitean, argitalpen zientifikoen kantitateak eta kalitateak garrantzi handia dute, eta, hori dela eta, argitalpenei buruzko hainbat datu eskatzen dira. Datu horiek ere alda daitezke proiektu, egiaztapen eta jardueraren ebaluazioari buruzko deialdiaren arabera, baina, oro har, garrantzitsuenak *kalitate-adierazleak* izenekoak dira. Oro har, gehien erabiltzen diren adierazleak bi dira: *eragin-faktorea* (argitalpenaren urtean aldizkariari egindako erreferentzia kopuruan oinarritutakoa), eta artikuluari egindako aipamen kopurua. Hiru eragin-faktore nagusi daude (JCR, Citescore eta SJR), eta bakoitza web-aplikazio jakin baten bidez kontsultatu behar da. Bestalde, artikuluari egindako aipamen kopurua erabilitako datu-basearen araberakoa da, eta zenbaki hori etengabe eguneratu behar da, ez baitzaio denbora-mugarik ezartzen.

Testuinguru horretan, argitalpen zientifikoei eta patenteei buruzko informazioa bilatzen eta biltzen duen tresna garatu da: *Xerka Online*. Aplikazioaren emaitza erabiltzailearen ikerketa-jarduerari dagokion CVaren zati handia da, eta, ondorioz, CVaren mantentze-lana erraztuko da eta denboraren aurrezpen nabarmena ekarriko dio irakasle/ikertzaileari. Ikertzaileak ataza horretarako behar duen denbora estimatzea zaila da, argitalpen kopuruaren araberakoa da. Hala ere, artikulu honen egileon kalkuluen arabera, irakasle/ikertzaile bakoitzak ataza honetan urtero 10 ordu inguru pasatzen dugu geure CVa eguneratzen eta mantentzen. *Xerka Online* tresnarekin, ordea, 30 segundo inguru besterik ez dira behar informazio eguneratuta lortzeko.

Tresnaren oinarrizko funtzionamendua 1. irudian adierazi da eskematikoki. *Xerka Online*k, web-scraping teknika erabiliz, hainbat datu-baseren web bidezko bilatzaileak erabiltzen ditu argitalpenak eta patenteak bilatzeko. Alde batetik, erabiltzailearen argitalpenak datu-base garrantzitsuenetan (Web of Science, Scopus, Crossref) eta UPV/EHUko Argitalpen Zerbitzuan bilatzen ditu. Ondoren, argitalpen bakoitzari dagozkion kalitate-adierazleak (JCR, Citescore eta SJR eragin-faktoreak eta aipamen kopurua) automatikoki lotzen dizkio (argitalpenak kalitate-adierazleak baldin baditu). Azkenik, erabiltzailearen patenteak Espacenet datu-basean bilatzen ditu.



1. irudia. Xerka Online aplikazioaren arkitektura.

*Xerka Online*n erabilerak denbora-murrizketa izugarria ahalbidetzen dio irakasle/ikertzaileari. Kalkulatu da argitalpen baten JCR, SJR eta CiteScore eragin-faktoreak (argitalpenari dagokion urtean) bilatzeko, aldizkariak arlo egokienean betetzen duen postua, tertzila (T1, T2, T3) eta koartila (Q1, Q2, Q3, Q4), eta aipamen kopurua bilatzeko 10 minutu behar direla. Proposatutako tresnarekin egilearen argitalpen guztiei buruzko informazioaren bilaketa segundo gutxian egiten da, eta, beraz, denboraren aurrezpena oso nabaria da. Jakina, zenbat eta argitalpen eta patente gehiago izan, orduan eta handiagoa izango da abantaila.

3. *MINERVA*

UPV/EHUko ikastegi bakoitzak horretan erabiliko den Kalitatea Bermatzeko Barne Sistema (KBBS) definitzen du, ikastegian eskaintzen diren titulazioen jarraipena egin ahal izateko. Kalitatea bermatzeko edozein sistematan bezala, KBBSaren helburu nagusietarikoa titulazioen hobekuntza da.

Gaur egun Vitoria-Gasteizko Ingeniaritza Eskolan eskaintzen diren ikasketa ofizialetan aplikatzen den KBBSa bat dator AUDIT Unibertsitate-prestakuntzaren Kalitatea Bermatzeko Sistemen Ezarpena Egiaztatzeko Programaren barruan indarrean dauden agirietan ezarritako arau eta ildoekin (ziurtagiria 2022ra arte dago indarrean). Sistema horretan, beste ba-

tzuen artean, honako kalitate-txosten hauek betetzen dira urtean zehar, barne-ebidentzia gisa eta urte amaierako txostena osatzeko erabiltzen direnak: i) irakasgai-txostenak (irakasgaiko koordinatzaileak betetzen ditu irakasgaiaren ohiko eta ezohiko deialdien ostean), ii) kurtso-txostenak (kurtsoko koordinatzaileak betetzen ditu) eta iii) gradu-txostenak (graduko koordinatzaileak betetzen ditu).

Txosten horietan, beste datu batzuen artean, irakasgai bakoitza gainditu duten ikasleen portzentajea biltzen da, matrikulatutako ikasleekin (gaindituak/matrikulatuak) eta azterketara aurkeztu diren ikasleekin (gaindituak/aurkeztuak) alderatuta. Beraz, datu horiek gradu-txostenean bildu ahal izateko, informazioa sortzeko prozesua honako urrats hauetan laburbil daiteke.

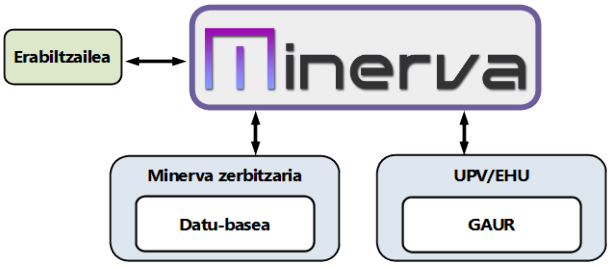
— Lehenengo pausuan irakasgai-txostenak sortzen dira. Horretarako, irakasgai bakoitzaren koordinatzaileak irakasgaiaren irakasle guztiei ohiko eta ezohiko deialdien emaitzen estatistikak eskatzen dizkie. Estatistikak kalkulatzeko, irakasle bakoitzak GAUR web-aplikazioan emaitzak kontsultatu behar ditu, eta estatistikak kalkulatu. GAUR-ek hainbat estatistika kalkulatzen ditu (gure eskolan eskatzen diren guztiak ez, ordea), baina aktak itxi baino lehen bakarrik kontsulta daitezke. Irakasle guztien emaitzak jaso ondoren, irakasgaiaren koordinatzaileak denak batuz kalkulatzen ditu irakasgaiaren estatistika orokorrak, eta irakasgai-txostenean sartzen ditu. Datu horietaz aparte, beste hainbat datu bete behar ditu eskuz *.doc* dokumentuan, eta, azkenik, e-postaz kurtsoko koordinatzaileari bidali.
— Ondoren, kurtso-txostenak sortzen dira. Kurtsoko koordinatzaileak koordinatzen duen kurtsoari dagozkion irakasgai-txosten guztiak jaso, eta estatistika eta datu guztiak bigarren *.doc* dokumentuan biltzen ditu. Dokumentu hori (kurtso-txostena) graduko koordinatzaileari e-postaz bidaltzen dio.
— Azkenik, gradu-txostenak kurtso-txostenak bezala sortzen dira: graduaren kurtso-txosten guztiak azken dokumentu batean batzen dira.

Prozesua ebidentziak izateko oso aproposa da, baina, aitzitik, hainbat desabantaila ditu. Alde batetik, irakasle guztiak GAUR web-aplikazioan hainbat aldiz sartu behar dira; bestetik, dokumentuak batzeko kopiatu-itsatsi eragiketa asko egin behar dira, eta akatsak egiteko probabilitatea handitu egiten da; eta, azkenik, e-mezu asko bidali behar dira. Oro har, prozesua ez da oso eraginkorra eta lan gehiena automatiza daiteke.

Minerva proiektuak sistema informatiko berria eraiki du GAUR web-aplikazioaren gainean, bete beharreko txostenen sorrera eta haien kudeaketa errazteko. Prozesu berria mahai gainerako aplikazio baten inguruan

diseinatu da (***2***). Kasu honetan, txostenak sortzeko prozesua honako hau da:

— Irakasgai-txostenak sortzeko, irakasleak aplikazioan ***Lightweight Directory Access Protocol*** (LDAP) kredentzialak sartuko ditu, eta programak automatikoki, web-scraping bidez, itxitako akten estatistikak kalkulatu eta *Minerva* zerbitzarian gordeko ditu. *Minerva* aplikazioaren bidez irakasgaiaren koordinatzaileek irakasgai-txostena beteko dute, eta estatistikak automatikoki itsatsiko dira txostenean.
— Kurtso bakoitzeko irakasgai-txosten guztiak bete direnean, kurtsoko koordinatzaileek beren kurtso-txostena beteko dute. Irakasgai-txosten guztiak dokumentu bakarrean kopiatuko dira, abantaila garrantzitsu batekin: kopiatu-itsatsi eragiketa automatikoki egingo da; horrela, irakasgai-txostenetan irakasgaiaren koordinatzaileek idatzitakoa kurtsoko koordinatzaileari automatikoki azalduko zaio, eta modu errazean bere balorazioa idatzi ahal izango du.
— Azkenik, gradu-txostenak betetzeko, graduaren kurtsoko koordinatzaileek betetako txostenak automatikoki azalduko zaizkio graduko koordinatzaileari, kurtsoari buruzko hausnarketa egin dezan.



2. irudia. Minerva aplikazioaren arkitektura.

*Minerva*ren abantaila nagusia aurreko sistemarekin konparatuz denbora-murrizketa da: irakasle bakoitzak bere irakasgaien estatistikak eskuz kalkulatzeko, 15 minutu inguru behar ditu. Aplikazioarekin, berriz, sarbidea askoz azkarragoa da, eta 30 segundo inguru behar ditu aktak jaitsi, estatistikak kalkulatu eta zerbitzarian gordetzeko. Beste abantaila bat da estatistiken kalkulu automatikoak akatsak murrizten dituela. Azkenik, aurreko prozedurarekin konpatibilitate arazo ugari izaten zituzten erabiltzaileek, dokumentu editore ezberdinak (besteak beste, Microsoft Word, Libre Office eta Open Office) eta bertsio ezberdinak erabiltzen zirelako. Oraingo sisteman, erabiltzaile guztiek aplikazio bera erabiltzen dute txostenak editatzeko, eta prozesuaren bukaeran dokumentu guztiak PDF formatura esporta daitezke klik bakar bat eginez.

Minerva aplikazioa 2019/2020 ikasturtean jarri da abian Vitoria-Gasteizko Ingeniaritza Eskolan, eta, nahiz eta oraindik goiz den ondorioak ateratzeko, 30 bat irakaslek beren zorionik beroenak helarazi dizkigute posta elektronikoaren bidez, sistema berriak aurrekoa hobetzen duelako.

4. **ONDORIOAK**

Artikulu honetan *web-scraping* teknikaren aplikazio berritzailea aurkeztu da: existitzen den sistema informatiko baten gainean sistema eraginkorragoa eraikitzea. Azpiko sistemaren programazioa aldatzeko aukerarik izan ez arren, teknika honek sistema zaharraren funtzionalitatea hobetzeko aukera ematen du.

Horrelako azpiegituretan oinarritutako bi tresna aurkeztu dira: *Xerka Online* eta *Minerva*. Bi proiektu horiek unibertsitateko irakasle/ikertzaileei zuzenduta daude. Batetik, *Xerka Online* aplikazioak CVaren mantentzelana murrizten du, eta, bestetik, *Minerva*k Vitoria-Gasteizko Ingeniaritza Eskolako KBBSaren baitan urtero sortzen diren txostenen sorrera eta kudeaketa automatizatu ditu.

Aurkeztutako aplikazioen abantaila nagusiak dira CVaren eguneratzean eta txostenen sorreran behar den denbora laburtzen dela eta prozesuan ager daitezkeen giza erroreak murrizten direla.

BIBLIOGRAFIA

[1] D. Gonzalez-Pena, A. Lourenco, H. Lopez-Fernandez, M. Reboiro-Jato eta F. Fdez-Riverola, «Web scraping technologies in an API world,» *Briefings in Bioinformatics,* 788-797 or., 2014.

[2] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso eta S. N. Mbaye, «Web Scraping: State-of-the-Art and Areas of Application,» 2019.

[3] L. Ulbricht, «Scraping the demos. Digitalization, web scraping and the democratic project,» *Democratization,* 426-442 or., 2020.

[4] E. Manica, C. F. Dorneles eta R. Galante, «Combining URL and HTML features for entity discovery in the web,» *ACM Transactions on the Web,* bol.13, 2019.

[5] M. A. Raza, B. Raza, T. Jabeen, S. Raza eta M. Abbas, «Using combined list hierarchy and headings of HTML documents for learning domain-specific ontology,» *International Journal of Advanced Computer Science and Applications,* bol.11, 233-239 or., 2020.

[6] J. C. Roldán, P. Jiménez eta R. Corchuelo, «On extracting data from tables that are encoded using HTML,» *Knowledge-Based Systems,* bol.190, 2020.

[7] J. Hillen, «Web scraping for food price research,» *British Food Journal,* bol.121, 3350-3361 or., 2019.
[8] J. I. Uriarte, G. R. Ramírez Muñoz De Toro eta J. M. C. Larrosa, «Web scraping based online consumer price index: The "IPC Online" case», *Journal of Economic and Social Measurement,* bol.44, 141-159 or., 2020.
[9] O. Jorge, A. Pons, J. Rius, C. Vintró, J. Mateo eta J. Vilaplana, «Increasing online shop revenues with web scraping: a case study for the wine sector», *British Food Journal,* 2020.
[10] H. Ahmed, T. A. Jilani, W. Haider, S. N. Hasany, M. A. Abbasi eta A. Masroor, «Producing standard rules for smart real estate property buying decisions based on web scraping technology and machine learning techniques,» *International Journal of Advanced Computer Science and Applications,* bol.11, 498-505 or., 2020.
[11] «MEFP», 2018. [Online]. Available: http://estadisticas.mecd.gob.es/EducaJaxiPx/Datos.htm?path=/Universitaria/Personal/EPU_2018-2019/PDI/Permanente//l0/&file=PDI0101.px&type=pcaxis. [Atzitze-data: 30 06 2020].
[12] D. Blazquez, J. Domenech eta J. Garcia-Alvarez-Coque, «Platforms impact with search data and web scraping,» *Measuring Technology,* 259-275 or., 2018.
[13] E. Uzun, «A Novel Web Scraping Approach Using the Additional Information Obtained from Web Pages,» *IEEE Access,* bol.8, 61726-61740 or., 2020.
[14] G. Naga Chandrika, S. Ramasubbareddy, K. Govinda eta E. Swetha, «Web Scraping for Unstructured Data Over Web», *Advances in Intelligent Systems and Computing,* bol.1076, 853-859 or., 2020.
[15] K. Mehta, M. Salvi, R. Dand, V. Makharia eta P. Natu, «A Comparative Study of Various Approaches to Adaptive Web Scraping,» *Lecture Notes in Electrical Engineering,* bol. 601, 1245-1256 or., 2020.