



Diana Teknologia Irtenbide berritzaileak

informazioa kudeatu eta
komunikatzeko

Information Retrieval
& Text Mining

DIANA TEKNOLOGIAK

XerKa aurkezten du, guk geuk sortutako irtenbide bat erabiltzaileak eskutatuko informazioa eraginkortasunez emateko.

Informazioa bilatu eta aztertzeke tresnek milioika dokumentu hainbat hizkuntzatan tratatzeko aukera ematen dute, honela dokumentu hauek ongi interpretatu eta sailkatzeko.

Berreskurapen sistema berriak informazioaren lokalizazio azkar eta eraginkor bat egiten du, hau dela eta, **XerKa** irtenbide onena da

Eduki Tematikoak aurkitu, kategorizatu eta berreskuratzeko

irtenbide eraginkor eta aurreratua

informazio sare digital
zabaletan

XerKa eskuragarri den edozein sare digital (Interneta, sare pribatuak, etab.) **etengabe ikertzeaz arduratzen den informazio sistema** da eta sistemaren erabiltzaileentzat interesgarri izan daitezkeen tokiak hautatzen ditu.

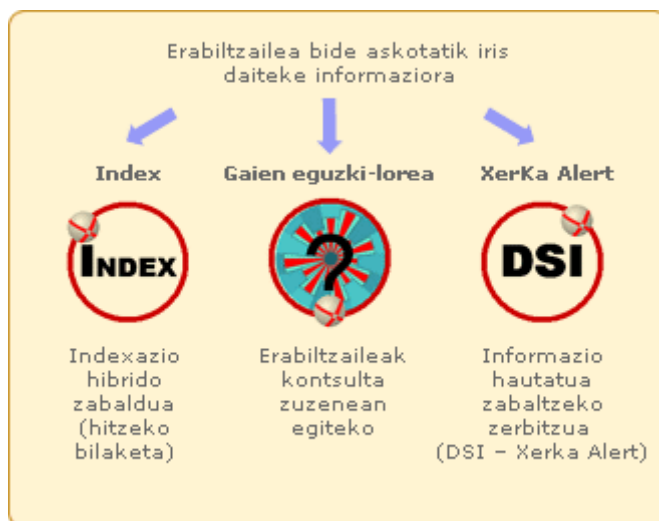
Zuzenean nahiz alerta sistema bitartez sartzeko aukera dagoen espazio tematiko multidimentsional bateko orri eta **dokumentuak iragazi eta sailkatzen ditu automatikoki.**

Erabiltzaileari informaziora azkar eta erraz iristeko aukera ematen dio, horretarako, berreskurapen modulu bat darabil; hona hemen modulu honek nola erantzuten duen:

- **azkar**: aurre indexazioa
- termino, kontzeptu edota gaiak bat datozelako: **multimodala**
- gaien arteko antzekotasunaren arabera: **multidimentsionala**
- gaztelaniaz, ingelesez, euskaraz eta frantsesez: **eleantza**

● teknologia eta merkatua
zaintzeko sistema gisa

● eduki tematikoak
modu inteligentean
sailkatzeko





Diana Teknologia Irtenbide berritzaileak

informazioa kudeatu eta
komunikatzeko

Puntu nagusiak

Zehaztasun eta
estaldura zabala
informazio
bilaketan

Bilaketa bideratua
(interesa pizten duten
tokietara)
Argumentuen
araberakoa
(gaiaren arabera)
Indartsua
(robot ugari)
Selektiboa
(interesgarria dena
iragazten du)

Hautatutako
edukien azterketa
sakona

Dokumentuaren
edukiaren azterketa
osoa.

Dokumentu motaren
arabera iragazia

eleanitza

Indexazio
multidimentsionala

Berrindexatze dinamikoa

Nabigazio erraza

Sarbide zuzena
dokumentura

Alerta zerbitzua

Lema eta hitzen
araberako bilaketa
osagarria

Indexazio multidimentsionala

XerKak dokumentuak hauen edukiaren arabera indexatzen ditu datu base multidimentsional batean. Hau da, dokumentu bakoitzak ardatz tematiko bakoitzarekiko lortzen duen puntuazioa adierazten du, honela, **irizpide askoren arabera geratzen da sailkatuta**

Honela, erabiltzaileak gai bat baino gehiago ukitzen dituen dokumentuak bilatu eta **lor ditzake**.

Erantzuna oso azkar jasotzen da, hau **XerKak** ezaugarri bat baino gehiago betetzen dituzten dokumentuak zuzenean berreskuratzen dituelako gertatzen da eta, beraz, betiko datu baseen moduan dokumentuak direktorio hierarkikoan arabera bilatzeko lan neketsua ekiditen da.

Eleanitza:
gaztelania, ingelesa, euskara eta frantsesa

XerKak dokumentuen eduki testuala aztertzen du lau hizkuntzatan eta egindako kontsulta bakoitzaren emaitzarako hautatutako dokumentua jatorrizko hizkuntzan aurkezten du. Erabiltzaileak sistemaren erantzuna nahi duzun hizkuntzara/hizkuntzetara mugatu dezake.

Lau hizkuntzatan koordinazioz lan egiten duenez, **XerKak** bilatzaileak terminoen arabera erabiltzean erabiltzaileak izaten duen arazoari egiten dio aurre: giltza-hitza egokia eta honen itzulpena beste hizkuntzetan zein den ez jakitea.

XerKak bitartez erabiltzaileak kontsulta bere hizkuntzan eta ardatz tematikoen arabera egiten du, beraz, ez du zertan giltza-hitzik erabili behar.

Aplikazioak

Teknologia eta/edo merkaturako zaintzeko sistema

Konpainiaren jarduerarako garrantzitsua den dokumentazioa bilatu, sailkatu eta bere esku jartzen dio.

Erabiltzaileak sortutako barne-informazio guztiari etekina ateratzea eta Interneten ateratako dokumentu berriak erabiltzea lortzen du.

Edukiaren gaiaren arabera sailkatzen du informazioa automatikoki

XerKak erabiltzaileari dokumentuak base batean sartzeko sailkapen irizpidea ezagutu behar izatearen lana kentzen dio

Basea berriz ere **indexatzen** uzten du, baita aldagai berriak sartzaren kasuan ere (ardatz tematikoak aldatu, gehitu edo ezabatu)

Formatu ugari

Word, Postscript, PDF, html, rtf,...

Euskarri diren plataformak

Windows 95/98/2000/Me/XP/NT
Linux, Solaris, AIX eta HP-UX.



XerKa Zainketa Sistema gisa erabiltzeko adibidea



Gaia:

empresa kudeaketa

XerKak sortutako ardatz tematikoei dagokien dokumentazioa bilatu eta sailkatzen du automatikoki: pertsonen kudeaketa, marketinga, informazioaren teknologiak, koiuntura ekonomikoa, finantzak, ...etab.

Rosa de los Temas



Gaien eguzki-lorea

Kontsulta:

Lan-taldeen antolaketari buruzko informazioa duen laneko segurtasun eta osasunari buruzko dokumentazioa.

Gaien eguzki-loreko interfaze grafikoan ardatz bakoitzaren luzerak gai horrek dokumentuetan izan behar duen garrantzia adierazten du. Erabiltzaileak kontsulta egiteko ardatz bakoitza luzatu edo moztu besterik ez du egin behar.

Adibidez:

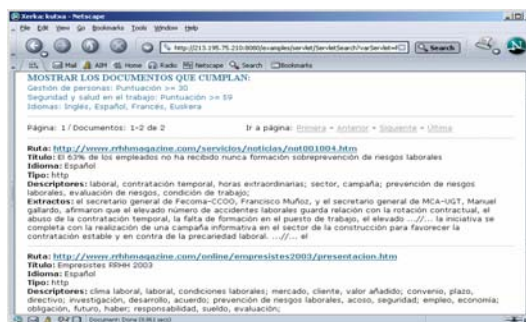
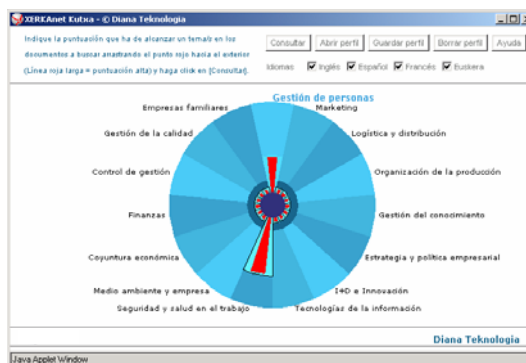
- Segurtasuna eta osasuna lanean 59 puntu
- Pertsonen kudeaketa 30 puntu

Erabiltzaileak kontsulta gorde dezake ("profilu sortu") hurrengo aldietan egiteko, bilaketaren emaitza nahi duen hizkuntzetara mugatu, ardatz berri bat luzatu galdera aberasteko, ...etab. Ardatz bat baino gehiago erabiltzeko aukerak aberasten du batez ere sistema, informazioa eremu jakin baten testuinguruan kokatzen baitu.

Erantzuna:

Sistemak erabiltzaileak eskatutako puntuazioa jaso duten **dokumentuen zerrenda** bat erakutsiko du. Zerrendak honako hau dakar: dokumentu bakoitzaren zati bat, dokumentua deskribatzen duten terminoen zerrenda bat eta kokatuta dagoen tokirako lotura.

Hautatutako dokumentuan klik egitea besterik ez dago **dokumentuaren jatorrizko bertsiora iristeko**.



XerKa Index

Index



Lemen araberako kontsulta osatzen du.

Adibidez: "administrazio-zehapen".

Sistemak "administrazio-zehapen" terminoa (terminoaren edozein forma) duten eta laneko segurtasun eta osasunarekin eta lan-taldean antolaketarekin zerikusia duten dokumentu guztiak erakutsiko ditu.

XerKa Alert

XerKa - Alert

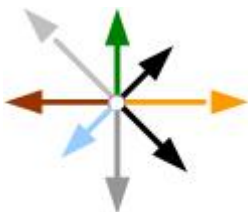


Informazio hautatua zabaltzeko programatuta dagoen sistema da.

Erabiltzaileak sortutako profilaren kontsulta egiten du egunero eta postaz **eskatutako edukia duten sartutako dokumentu berrien zatiak** bidaltzen dizkio.

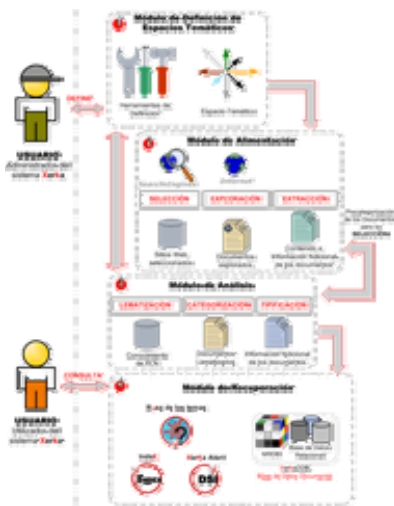
Moduluak

Zer informazio nahi dugu?



Gaiak

Nahi dugun informazioa aurkitzeko



Informazioa lortzeko modu asko

Rosa de los Temas XerKa Cluster



XerKa Index

XerKa Alert



Gaiak kudeatzeko modulua

1



Erabiltzailearen, lan-talde baten edo erakunde baten nahiak islatzen dituen gai multzoa definitzeaz eta elikatzeaz arduratzen da.

Ardatz tematikoak gai horri dagokion terminologia landuz zehazten dira. Hasierako termino multzo hauek hizkuntza teknikariek darabilten hitzen datu basean nahiz elikatze moduluko metabilatzaileetako argumentazio sisteman sartzen dira.

Bilaketaren emaitza "formen jarraipen terminologikora" bideratutako algoritmoen arabera prozesatzen da; algoritmo hauek terminoen garrantzia berriz baloratzeko, erabiltzen ez diren terminoak aurkitzeko eta termino berriak sortu direla aurkitzeko gai dira.

Sortzen diren gai eremu dinamiko hauek Xerkaren elikatze eta berreskurapen moduluen euskarri dira.

2

Elikatze modulua

Azterketa modulua prozesatu behar dituen dokumentuak bilatu eta lortzeaz arduratzen da.

Web-orriak bilatzeko Interneteko bilatzaileei zehaztutako gaietan oinarritzen diren argumentuak erabiliz egiten zaie kontsulta. Bilaketaren emaitza aurreanalizatu egiten da gai bakoitzaren barruan duen garrantzia zein den jakiteko. Bilaketa argumentuak doitu eta bilaketaren emaitza onenak ematen dituzten terminoak erabiltzen dira.

3

Azterketa Modulua

Azterketa ezberdinak egiten ditu, honela gai kudeaketaren elikatze sistema, berreskurapen moduluko indexazio sistema eta elikatze moduluko Web-orrietako balorazio prozesua elikatzen da.

Dokumentuetako eduki testuala lau hizkuntzatan aztertu eta esanahi unitateetara ekartzen da, lema sinpleetara, konplexuetara edo kontzeptuetara. Unitate hauek zehaztutako gaien testuinguruan erabili eta jarraipen terminologikoko algoritmoak eta kategorizazio algoritmoak elikatzen dira.

Dokumentuaren egituraren eta testuaren informazioa dokumentuak talde ezberdinetan tipifikatu edo sailkatzeko erabiltzen da: ikastaroak, curriculumak, glosategiak, eta abar.

4

Berreskurapen Modulua

Dokumentuak datu base erlazional eta multidimentsional batean (MRDB) indexatzeaz arduratzen da; datu base honetan dokumentuak modu ezberdinetan gordetzen dira ondoren berreskuratu ahal izateko.

Gaikako indizeen bide erlazionalak dokumentuak bilatzea errazten du bilaketa gai bakarraren arabera egiten denean. Bide multidimentsionalak, berriz, gaiaren barruan antzekotasunen arabera bilaketak egiten ditu eta gai batean baino gehiagotan aldi berean azaltzen diren dokumentuak ere bilatzen ditu. Azkenik, **lemen arabera sarbidearekin** Interneteko bilatzaileetan egiten den moduan bilaketak egin ditzakegu dokumentu basean (**Xerka - Index**).

Erabiltzailea modu interaktiboan sar daiteke **sarbide multidimentsionaleko interfaze grafikoaren** bitartez (**Gaien eguzki-lorea**) edo modu programatuan **alerta sistema** gisa (**Xerka Alert - DSI**)