

# Sistema recomendador de materiales educativos para plataforma LMS

Memoria del projecte final de grau

Ponent	Enrique Romero Merino
Departament	Computació
Director	Andrea Vega Martínez
Autor	Albert Carvajal Abellan

# Index

<b>Index</b>	<b>2</b>
<b>Agraïments</b>	<b>4</b>
<b>Resum</b>	<b>5</b>
<b>Resumen</b>	<b>6</b>
<b>Abstract</b>	<b>6</b>
<b>Introducció</b>	<b>6</b>
Empresa	6
Arquitectura de l'empresa	7
Objectiu	8
Estat de l'art	9
Recomanadors	9
Content-Based	9
Collaborative Filtering	9
Sistemes Híbrids	10
Abast	10
Stakeholders	11
<b>Metodologia</b>	<b>11</b>
Metodologia	11
Planificació inicial	11
Reunió de planificació	11
Investigació de mètodes actuals	12
Familiarització amb l'entorn del projecte	12
Preparació de l'infraestructura	12
Redisseny d'interfície	13
Anàlisi dels resultats	13

Optimització de còdi	13
Desenvolupament de tests	13
Planificació final	13
Definició de l'anàlisi funcional	14
Definició del model de dades relacional	14
Definició del model de dades Lucene	14
Implementació del motor recomanador	14
Interfície de frontend a backend	15
Implementació Disseny Frontend	15
Endpoint d'enllaç amb el material	15
Anàlisi d'alternatives	15
Mètodes de validació	16
Eines	16
Riscos	17
<b>Recursos</b>	<b>17</b>
Recursos	17
Pressupost	18
<b>Integració dels coneixements</b>	<b>19</b>
Anàlisi	19
Disseny	19
Model	21
Criteris	23
Aleatori	23
Criteri segons l'Expedient Acadèmic de l'Usuari	23
Criteri segons Similitud d'Usuaris	24
Criteri segons l'Usuari marca com a Favorit	24
Criteri segon valoració positiva de l'Usuari	24

Batch	25
Tests	25
<b>Sostenibilitat</b>	<b>26</b>
Matriu de sostenibilitat	26
Anàlisi ambiental	26
Consum de disseny	26
Empremta ecològica	26
Riscos ambientals	26
Anàlisi econòmic	27
Factura	27
Pla de viabilitat	27
Riscos econòmics	27
Anàlisi social	27
Impacte personal	27
Impacte social	27
Riscos socials	27
Lleis i regulacions aplicades	28
<b>Conclusions</b>	<b>28</b>
Resultats	28
Futur	28
Conclusions	29
Competències tècniques aplicades	29
<b>Glossari</b>	<b>30</b>
<b>Bibliografia</b>	<b>31</b>

# Agraïments

Gràcies a l'Andrea per donar-me la responsabilitat i la guia per dur a terme tot el projecte a nivell d'empresa, a l'Enrique per guiar-me en el procés a nivell acadèmic, a en Jordi per ensenyar-me i guiarme durant el projecte, a Robert per ajudar-me en tots els problemes en els que no hi veia sortida i sobretot a tots els companys de l'equip que m'han donat molt de recolzament en tot.

## Resum

L'empresa GEC (GESTIÓN DEL CONOCIMIENTO SA) ofereix solucions de formació i gestió del coneixement adaptades a les necessitats de grans empreses per millorar l'eficiència dels seus treballadors. Aquest projecte duu a terme el desenvolupament d'un sistema de recomanació de continguts acadèmics. A l'hora de generar recomanacions s'haurà de tenir en compte la interacció de l'usuari a la plataforma, així com el rendiment dels diferents algorismes aplicats. La solució desenvolupada haurà de ser conscient de les limitacions que la plataforma ofereix, tant d'arquitectura de software com de hardware.

## Resumen

La empresa GEC (GESTIÓN DEL CONOCIMIENTO SA) ofrece soluciones de formación y gestión del conocimiento adaptadas a las necesidades de grandes empresas para mejorar la eficiencia de sus trabajadores. Este proyecto lleva a cabo el desarrollo de un sistema de recomendación de contenidos académicos. A la hora de generar recomendaciones deberemos de tener en cuenta la interacción de los usuarios en la plataforma, así como la arquitectura de la plataforma, el rendimiento de los diferentes algoritmos aplicados. La solución desarrollada deberá ser consciente de las limitaciones que la plataforma ofrece, tanto de arquitectura de software como de hardware.

## Abstract

The company GEC (GESTIÓN DEL CONOCIMIENTO SA) offers formation solutions and knowledge management tailored to the needs of big companies to improve the efficiency of their employees. This project carries out the development of a recommendation system of the academic contents of the platform. At the time of generating the recommendations we will have to be aware of the interaction of the user, just like the performance of the algorithms applied. The developed solution should be aware of the limitations on the platform, such software architecture as hardware architecture.

# Introducció

## Empresa

El projecte que es tracta en aquesta memòria esdevé d'un treball de fi de grau de modalitat tipus B, i com a tal, s'ha realitzat en el context d'una empresa. Degut a que el contracte amb aquesta empresa s'ha vist finalitzat de manera prematura, molta de la informació que es tracta és confidencial, de manera que es veurà modificada i/o eliminada, per tal de preservar la confidencialitat estipulada per l'empresa.

L'empresa en qüestió es diu GEC (GESTIÓN DEL CONOCIMIENTO SA), i ofereix solucions de formació i gestió del coneixement adaptades a les necessitats de grans empreses per millorar l'eficiència dels seus treballadors. Per això, defineix estratègies de formació, produeix i imparteix cursos i plans de carrera acreditats en línia, desenvolupa i dinamitza comunitats virtuals de gestió del coneixement.

L'organització va néixer al 1996 com a societat amb l'objectiu d'identificar, analitzar, promoure i impulsar la extensió dels ensenyaments no presencials i de qualsevol altra forma de comunicació no presencial per tot el món mitjançant procediments i recursos propis, concertats i externs.

El 1997 es va registrar com a empresa privada amb la UOC (Universitat Oberta de Catalunya) i La Caixa com a màxims inversors, i posteriorment, a l'any 2009 va ser venuda, amb l'aprovació de la Generalitat de Catalunya, a dos fons d'inversió. Finalment, entre l'any 2012 i 2013, la consultora especialitzada en formació permanent i desenvolupament de persones Grupo Hedima va comprar-ne el 100% de la mateixa.

GEC està formada per més de 100 professionals, amb una facturació de més de 4 milions d'euros anuals, amb presència a més de 40 països i amb grans marques com les que apareixen a la següent figura.



## Arquitectura de l'empresa

Les plataformes principals proporcionades per l'empresa per donar suport al sistema LMS (Learning Management System) és una plataforma en un entorn web, i en una aplicació mòvil.

Per posar una mica en context, de les aplicacions esmentades, aquestes estan presents a 43 països, disponibles en 13 idiomes, han estat utilitzades per uns 1.6 milions d'usuaris i en total es poden sumar gairebé 22 milions d'hores d'aprenentatge impartides per aquestes aplicacions.

La mateixa empresa ofereix: allotjament, manteniment i desenvolupament de millores evolutives de les aplicacions, acotades a les peticions de cada client.

Degut a que tots els projectes de cada client utilitzen com a base Virtagora, cada vegada que s'actualitza Virtagora en una nova versió, tots els projectes es veuen afectats per aquests canvis.

Per fer els desenvolupaments de noves característiques al producte base, s'utilitza una aplicació plantilla, anomenada Virtagora Factory (VF). És una aplicació client, però buida de cap tipus de codi personalitzat. Aquesta aplicació és la que s'utilitza per fer el desenvolupament de nous components comuns, i on ens hem d'assegurar que funciona correctament.

Per la gestió d'incidències i nous desenvolupaments s'utilitza l'eina de gestió Jira. Aquesta plataforma és la que s'utilitza com a canal de comunicació entre els diferents actors involucrats al desenvolupament de la plataforma: clients, administradors, project managers, i desenvolupadors.

Tots els projectes consten de quatre entorns: Entorn local, Integració (INT), Pre Producció (PRE) i Producció (PRO). L'entorn local és on cada desenvolupador desenvolupa i testeja el codi. INT és l'entorn utilitzat pels desenvolupadors per testejar l'integració del codi amb els demés components en desenvolupament. L'entorn de PRE és on client pot realitzar les seves proves i donar el vistiplau dels canvis que s'han desenvolupat. PRO és l'entorn que client fa servir per treballar i on l'usuari final hi accedeix.

La gestió de codi de la plataforma i dels diferents projectes de cada client es duu a terme mitjançant GitLab. La gestió de branques que s'utilitza a als projectes dels clients és la de *feature branching*. Cada aplicació disposa de un repositori per a ell sol. A cada repositori hi trobarem una branca per cada entorn: INT, PRE i PRO. Cada desenvolupament que es realitzi a un projecte, ha d'estar associat a un tiquet de Jira.

El sistema LMS, és un sistema compost per diferents elements i tecnologies. El llenguatge de programació utilitzat pel desenvolupament és Java, amb els frameworks Hibernate i Struts. Per l'emmagatzematge de dades s'utilitza principalment una base de dades Oracle, tot i que també existeix una base de dades MySQL on s'emmagatzema les estadístiques generades a la plataforma, i



les dades indexades pel motor de cerca Lucene, són emmagatzemades en el sistema de fitxers linux del servidor on s'hi troba la plataforma. La plataforma també ofereix altres sistemes que s'han integrat amb diferents tecnologies, com el modul de BI que ofereix Yellowfin, el gestor de vídeos de Brightcove, o el servei de gestió de correus electrònics de Zimbra. A l'hora de la gestió de dependències i construcció del projecte, s'utilitza el gestor Gradle, que compila tant el projecte base(Virtagora), com descarrega els paquets de codi del repositori de dependències, i les inclou al projecte client per després generar el projecte compilat que sera desplegat al servidor Tomcat pertinent.

## Objectiu

L'objectiu del projecte és poder oferir als clients, un producte que millori l'experiència d'usuari, i en conseqüència, augmenta el temps que passa a la plataforma. Aquest objectiu ve degut a que tots els continguts de la plataforma són educatius, de manera que un major temps a la plataforma vol dir que l'usuari s'ha exposat a més formació que la estrictament exigida per l'empresa contractant. Això implica una ampliació de competències i/o coneixements que influiran directa o indirectament al seu rendiment al seu lloc de treball.

Per tal d'assolir aquest objectiu, el producte per la qual s'ha decidit, és un mòdul recomanador de materials acadèmics. Oferint continguts acadèmics acotats als interessos de l'usuari, augmenta les probabilitats de que un usuari hi accedeixi, fora de les obligacions imposades per l'empresa contractadora.

## Estat de l'art

### Recomanadors

Degut a que a la nostra plataforma no existeix un sistema d'aquestes característiques, a banda del mòdul que hem esmentat a l'apartat anterior, que tot i que la intenció sigui la mateixa, no utilitza cap sistema automatitzat. Els sistemes de recomanació són una eina derivada de la necessitat d'exploració de les dades. Degut a que cada vegada més ens trobem amb sistemes massius de dades, és requisit indispensable l'ús d'eines que ens ajudin a la tasca de filtrar aquesta informació. Un dels sistemes que van sorgir com a solució a aquest problema, van ser els sistemes recomanadors. Hi ha dos tipus bàsics de sistemes recomanadors: Els sistemes basats en el contingut, i els sistemes amb filtres col·laboratius. Actualment el sistema que més s'utilitza és un que utilitza alguna combinació d'aquests dos sistemes. Per exemple, Amazon utilitza una combinació d'aquests dos de manera que et recomana productes similars a aquells que has comprat, combinat amb els articles amb els que usuaris com tu han comprat.

## Content-Based

Aquests tipus de sistemes utilitzen la informació inherent a cada entitat. Aquest tipus de mètodes utilitzen les recomanacions com si fossin problemes de classificació per a cada usuari. En aquests sistemes normalment s'utilitzen paraules claus per descriure els elements i l'usuari es veu forçat a indicar el tipus d'elements que l'hi agrada.

## Collaborative Filtering

Aquest tipus de sistemes es basen en la hipòtesi de que persones que han coincidit en certes preferències anteriorment, coincidiran en el futur. Aquests sistemes generen recomanacions utilitzant les accions dels usuaris com a mètode de valoració de l'entitat. Trobant altres usuaris que han realitzat accions similars, generarem recomanacions utilitzant aquest tipus de veinari.

Aquests tipus de sistemes solen tenir 2 tipus de problemes associats als seus mètodes de recomanació:

- Cold start: Per a nous usuaris o entitats, no hi ha informació suficient per generar recomanacions adients.
- Escalabilitat: En molts dels sistemes on es troben aquests sistemes hi ha milions d'usuaris i entitats, de manera que una gran potència de càlcul és requerit per a generar aquestes recomanacions

## Sistemes Híbrids

Els sistemes híbrids són els sistemes resultants de la combinació de la resta. Aquests sistemes pretenen minimitzar els punts febles d'un sol sistema, convintant-los amb d'altres i així maximitzar els beneficis. S'ha d'anar amb compte de que la combinació dels sistemes no potenciï els punts febles de cada un i com a resultat sigui un sistema pitjor que qualsevol de les seves parts. Per exemple, aquests sistemes es solen utilitzar per intentar mitigar el problema de l'arrencada. En el cas dels sistemes de Collaborative Filtering ens trobem que en el cas inicial on no hi ha dades, ni interacció entre elles, el sistema es veu mancat d'informació com per generar cap resultat. En canvi, si aquest sistema es combina amb un sistema Content-Based, obtindrem que, tot i no tenir interacció entre elements, podem oferir algun resultat.

Alguns dels sistemes utilitzats per a la hibridació són:

- Ponderació: Es combina la puntuació de diferents criteris de recomanació, numèricament.
- Substitució: Es tria un criteri de recomanació sobre la resta i s'aplica.
- Combinació: Es generen les recomanacions per a diferents sistemes i es mostren totes.

- Combinació d'atributs: Atributs derivats de diferents fonts de coneixement son combinats i s'alimenta a un sol algoritme de recomanació.
- Augment d'atributs: Es computa un atribut o conjunt d'atributs, el qual és després part de l'entrada de la següent tecnica.
- Meta-level: Una tecnica de recomanació s'aplica i produeix un model el qual servirà d'entrada per a la següent tecnica.

## Abast

Aquest projecte es centra en l'ús de les dades de la base de dades d'Integració del projecte de VF, tot i que aquesta base de dades no aportí informació sobre la qualitat de les recomanacions, si ens aportarà informació sobre el rendiment i l'impacte que tindrà el mòdul a la plataforma en el sentit de rendiment. S'incorporarà el model de dades necessari per implementar el framework per a generar les recomanacions i s'implementaran diferents criteris que generin recomanacions per tal de ponderar-les enriquint així les recomanacions.

Molts sistemes utilitzen sistemes d'interpelació amb l'usuari per tal d'extreure preferències sobre les diferents categories d'entitats. Haurem de tenir en compte això per preparar-ho per a futurs evolutius que es vulguin implementar al respecte.

## Stakeholders

L'equip de persones que es veurà involucrada en el desenvolupament del projecte és:

- Andrea Vega: Directora del TFG
- Jordi Roca: Responsable de producte.
- Fran Collado: Encarregat de la programació de frontend.
- Alex Morales: Encarregat de la programació de backend.
- Albert Masedo: Encarregat de programació de backend.
- Albert Carvajal: Encarregat de dur a terme el projecte i documentar-lo

Els actors més interessats del projecte són els clients, ja que aquest projecte aporta un gran valor a la plataforma i un gran impacte a les seves empreses. La propia empresa és un dels actors més interessats del projecte, ja que, com he dit abans, aquest projecte aporta un gran valor a producte, i els fa més competitiu davant de la competència.

# Metodologia

## Metodologia

Seguidament veurem la metodologia planificada en contraposició de la metodologia aplicada finalment, i com s'ha vist afectada pel context on es desenvolupa. Degut a que l'entorn de desenvolupament és el d'una empresa, hi ha moltes parts que hi intervenen i que poden afectar i/o alterar als processos del projecte.

## Planificació inicial

### Reunió de planificació

La planificació inicial va ser resultat de les diferents reunions que vam tenir amb el Director de tecnologia, l'encarregat de Disseny i el Cap de projecte. Les diferents reunions van donar com a resultat un diagrama de Gantt i un anàlisi de requeriments que s'hauran de respectar per tal d'assolir l'èxit del projecte.

El temps d'aquesta tasca és de: 5h

### Investigació de mètodes actuals

Desde inicis de gener s'ha recopilat informació existent dels diferents mètodes utilitzats a la recomanació d'entitats. Es va començar buscant documentació acadèmica, com Introduction to Recommender Systems Handbook. Una vegada es va tenir una visió global del món acadèmic en relació als sistemes recomanadors, es va donar un cop d'ull al món comercial. Amb el propòsit d'oferir una visió comercial sobre el producte, es va investigar alguns documents d'algunes de les empreses més importants, com el sistema d'Amazon i el seu sistema item-to-item collaborative filtering. Aquesta tasca va requerir de recursos humans necessaris per adquirir aquest coneixement, a banda d'un ordinador per investigar la documentació necessària.

El temps d'aquesta tasca és de: 40h

### Familiarització amb l'entorn del projecte

Degut a que aquest projecte ha de ser incorporat a una plataforma ja existent, s'ha de donar una part previa de coneixement de la plataforma, per a poder avaluar i analitzar els conceptes més rellevants a tenir en compte per a les recomanacions. S'ha d'anar amb compte, degut a que qualsevol canvi que es realitzi afectara a 22 projectes clients, de manera que haurem de ser molt curosos amb els canvis que introduïm. En aquest procés s'ha realitzat amb la col·laboració d'alguns desenvolupadors senior amb una llarga experiència amb la plataforma. Al llarg d'aquesta tasca s'ha arribat a la conclusió de que la millor

solució serà aprofitar el motor de cerca Lucene, que ja és incorporat a dins de producte. Aquesta tasca ha requerit dels recursos humans prèviament esmentats.

El temps d'aquesta tasca és de: 40h

## Preparació de l'infraestructura

Abans de re, s'ha d'adequar el context a la arribada del nou projecte. El qual vol dir que s'ha d'introduir algunes modificacions a les dades que es guarden a Lucene, com les relacions que hi ha entre usuaris i entitats. S'haurà de realitzar el desenvolupament de l'infraestructura on s'executaran els calculs dels diferents criteris, i les ponderacions. Aquest sistema haurà de ser fàcilment modulable, i ampliable. Donat a que hi ha 22 clients diferents amb necessitats diferents cada un, es pugui personalitzar cada criteri de manera que tot segueixi funcionant. Els recursos necessaris per aquesta tasca hauran de ser: El codi del projecte, servidors, bases de dades i els índexs generats per Lucene. A més es requeriran dels recursos humans per a especificar i desenvolupar els canvis evolutius pertinents.

El temps estimat d'aquesta tasca és de: 120h

## Redisseny d'interfície

Esta parte del proyecto también se deberá tener en cuenta, aunque los encargados de dicho desarrollo, no sean los autores del proyecto.

Esta tarea deberá de realizarse juntamente con el desarrollo de la infraestructura para poder encajar el módulo en el contexto de la plataforma. Deberá hacerse, coordinando la sección de diseño y desarrollo frontend de la empresa. Habrá dependencia entre las tareas de rediseño y desarrollo de la infraestructura, ya que para poder visualizar los resultados, se deberán mostrar en algún sitio.

El temps estimat d'aquesta tasca és de: 40h

## Anàlisi dels resultats

Després del primer desenvolupament, és necessari una etapa d'anàlisi de resultats. Aquests anàlisis hauran d'incloure la qualitat de les recomanacions i rendiment de la plataforma, sota diferents escenaris (és a dir, tests d'estrès). Amb les dades dels diferents tests, haurem d'extreure els punts de conflicte, i haurem d'optimitzar i realitzar els canvis de codi que siguin pertinents.

El temps estimat d'aquesta tasca és de: 8h

## Optimització de còdi

Una vegada haguem realitzat l'anàlisi i extret els punts crítics, haurem d'aplicar optimitzacions sobre el codi, com reduir la dimensionalitat de les dades

emmagatzemades. O optimitzar els mètodes de recomanació, com els que va introduir Amazon al seu codi.

El temps estimat d'aquesta tasca és de: 40h

## Desenvolupament de tests

Per a la correcta integració del mòdul a la plataforma, i als posteriors clients, s'ha d'implementar una serie de tests, per assegurar el correcte funcionament, tant del mòdul, com de l'integració d'aquest a la plataforma.

El temps estimat d'aquesta tasca és de: 40h

El temps estimat del projecte és de: 333h dels quals 293h van a carrec meu.

## Planificació final

La primera part de la planificació inicial va transcórrer tal com va ser planejada, ja que quan es va fer la planificació ja s'havien realitzat dites tasques. A partir d'aleshores les tasques es van anar alterant i modificant, creant-ne de noves i modificant de velles.

## Definició de l'anàlisi funcional

Aquesta tasca va sorgir arrel de les polítiques d'empresa que obliguen a que tot desenvolupament hagi de ser documentat funcionalment prèviament al desenvolupament. Aquesta tasca va introduir una complexitat que no s'havia previst inicialment, ja que es va requerir d'una formació adicional sobre les documentacions funcionals. Aquesta manca d'experiència amb la matèria, va produir que es fessin quatre entregues d'aquesta documentació amb diferents afegits a cada una. Dins a l'anàlisi funcional final es va incloure una descripció dels mètodes que s'aplicaran, un exemple de com funciona l'algorisme que genera les recomanacions, i un glossari de termes que s'utilitzen en el projecte.

Degut a temes de confidencialitat, aquestes addicions i el contingut exacte de la documentació final no es pot mencionar de cap manera.

El temps d'aquesta tasca és de: 160h

## Definició del model de dades relacional

Aquesta tasca és una subtasca esdevinguda de la tasca de [preparació de l'infraestructura](#). Es va convertir en una tasca necessària a petició del responsable de producte. Ja que les multiples restriccions que aporta la plataforma, les quals en parlarem més endavant, ens força a que el model de dades hagi de tenir una estructura concreta. La realització d'aquesta tasca va ser

molt més àgil, ja que una vegada ens van plantejar les restriccions, l'estructura va ser molt més clara.

El temps d'aquesta tasca és de: 40h

## Definició del model de dades Lucene

Aquesta tasca, com l'anterior, ha sigut esdevinguda de la tasca de [Preparació de l'Infraestructura](#). Com l'anterior també va ser una tasca necessària a petició del responsable de producte. Dins d'aquesta tasca es va introduir el fet de desenvolupar nous models de dades per lucene a fi de no interferir amb el model de dades que lucene ja estava recollint. Això ens va presentar una gran oportunitat de realitzar un model mantenible i ampliable.

El temps d'aquesta tasca és de 8h

## Implementació del motor recomanador

Aquesta tasca és la tasca principal en la que jo em veure personalment involucrat. Sota aquesta tasca hi penja tota la implementació dels diferents dissenys prèviament esmentats i l'implementació del motor recomanador, així com els processos nocturns que realitzaran. S'han de tenir en compte que existeixen requeriments i limitacions que s'han de complir. Aquestes limitacions s'introduiran a l'apartat d'[Integració dels coneixements](#).

El temps d'aquesta tasca és de 240h

## Interfície de frontend a backend

Aquesta tasca és una subtasca derivada de la tasca de Preparació de l'Infraestructura. I es una tasca generada per a introduir a altres desenvolupadors per agilitzar la producció del projecte. Aquesta tasca ha de resoldre la part de backend que connecta les recomanacions generades pel motor, amb la part de frontend.

El temps d'aquesta tasca és de 60h

## Implementació Disseny Frontend

Aquesta tasca ha sigut desenvolupada per un membre de l'equip de frontend. Aquesta tasca és derivada de [Redisseny d'Interfície](#).

El temps d'aquesta tasca és de 80h

## Endpoint d'enllaç amb el material

Aquesta tasca és una subtasca derivada de la tasca de [Preparació de l'Infraestructura](#). I es una tasca generada per a introduir a altres

desenvolupadors per agilitzar la producció del projecte. Aquesta tasca ha de resoldre la part d'enllaçar els elements mostrats per frontend cap a la pagina del recurs en qüestió.

El temps d'aquesta tasca és de 16h

Temps total de desenvolupament del projecte: 604h

## Anàlisis d'alternatives

Com a primera alternativa sempre hem de considerar el fet de realitzar els càlculs de forma manual. Després de presentar el cas base, anem a mirar altres alternatives. Si cerquem per internet, les primeres alternatives amb les que ens toparem serán les plataformes de cerca Elasticsearch i Apache Solr.

Elasticsearch és un projecte basat en el motor de Lucene distribuït i amb una interfície web RESTful i amb documents JSON. Aquesta eina te grans avantatges, ja que esta desenvolupat en Java, el que vol dir que és compatible amb la nostra plataforma. Té una gran velocitat de resposta. I es distribuït, el que fa que sigui fácilmente escalable i també el converteix en altament compatible amb la nostra plataforma, ja que tenim entorns amb múltiples instàncies de l'aplicatiu. El projecte és de codi obert sota les condicions de la llicència Apache.

Apache Solr és un projecte basat en el motor de Lucene. Corre sobre contenidors de servlets Java, el que el faria perfecte per al nostre aplicatiu.

El major problema d'aquests projectes és que, implementar una nova plataforma de cerca de dades és totalment inviable, ja que l'abast d'això és un projecte en si mateix. Sense contar amb la migració dels canvis a cada un dels 22 projectes de client. Com el factor comú de les alternatives, és el mateix motor que ja tenim integrat, que és Lucene, s'ha triat d'un punt intermig entre la primera alternativa i la resta.

## Mètodes de validació

Les validacions que es requereixen fer del desenvolupament per garantir el correcte funcionament vindran de part dels Tests. Els tests han de garantir el correcte funcionament a nivell atòmic de cada funció implementada, així com també, ha de garantir el correcte funcionament de l'integració de totes les funcions entre si, i que la funcionalitat s'hagi programat correctament. L'explicació dels tests desenvolupats s'ampliarà a la secció de [Tests](#). La suite que s'utilitzarà a l'hora dels tests es la de JUnit, amb l'eina Mockito per moquejar les classes i mètodes que toqui a l'hora de fer tests unitaris. En el cas dels tests d'estrès les proves es faran sobre la plataforma directament a l'entorn de PRE del projecte de VF.



## Eines

Les eines involucrades en el projecte són les següents:

- SourceTree: Utilitzat per a la gestió del repositori
- IDE Eclipse Luna: Utilitzat per a desenvolupar el codi del projecte.
- Superputty: Utilitzat per a accedir al servidor i fer control i manteniment dels logs del aplicatiu.
- JIRA: Utilitzat per a la gestió de les tasques i incidències.
- Navegador Chrome: Utilitzat tant per a accedir al servidor per fer-hi proves, com per buscar documentació.
- Repositori de codi GitLab: Utilitzat per a mantenir el control de versions del codi
- SQL Developer: Utilitzat per accedir a la base de dades
- Oracle: Servidor de Base de dades
- Apache Tomcat: Utilitzat per arrancar els projectes
- GanttProject: Utilitzat per a la planificació de les tasques
- Enterprise Architect: Utilitzat pel disseny del model de dades
- Windows: Utilitzat als ordinadors dels desenvolupadors
- Linux: Utilitzat en els servidors per hostejar tots els projectes
- Microsoft Office: Utilitzat per a la realització de la documentació

## Riscos

Aquest projecte, si no es complís la restricció de que no es pot tocar cap classe ja existent, seria un projecte de molt alt risc. Aquest risc ve degut a aquest codi és a 22 projectes, on cada projecte molt probablement tingui parts personalitzades, que es poden veure afectades si es realitza qualsevol canvi a producte. Això afegit a que el codi és molt ampli i complex, amb molts components interactuant entre si a la vegada, provoca que la millor solució sigui desenvolupar tots els components com a nous elements del projecte. Amb nous elements podem estar segurs que no hi haurà codi personalitzat implicat a cap projecte, i que tots els projectes clients rebran una versió funcional del recomanador. Només amb aquesta restricció el projecte és viable, d'altra manera el desenvolupament seria massa costós com per poder assumir els riscos.

Un dels riscos del projecte, també és que es faci una migració de les llibreries de Lucene a una versió més actual i deixi de funcionar el mòdul recomanador.

Aquest risc és menor, ja que juntament amb cada versió que Lucene treu, s'adjunta un document que explica quines funcions s'han modificat, i quins passos s'ha de seguir per a realitzar la migració.

Un altre dels riscos del projecte és el de que es quedi a mitges. Per a aquest problema s'ha dissenyat el projecte de manera, que sigui modular, i el desenvolupament pugui realitzar-se en diferents etapes. De manera que aquest risc és un risc mitjà, ja que un mal disseny, podria comportar que el projecte quedi totalment inservible en algun punt.

## Recursos

### Recursos

Els recursos invertits en aquest projecte han sigut recursos materials i recursos humans.

Pel que fa pels recursos humans:

- 4 Desenvolupadors
  - 3 Desenvolupadors de Backend
  - 1 Desenvolupador de Frontend
- 1 Cap de Projecte
- 1 Dissenyador
- 1 Director de tecnologia

Recursos materials:

- 4 ordinadors de sobretaula
- Material d'oficina
- Impresora
- Servidor d'integració
- Servidor de Reproducció

Recursos energètics:

Tots els programadors involucrats treballen amb ordinadors fixes d'alt rendiment amb un consum d'uns 150W. Com tenim 604h de treball, el càlcul aproximat de

consum energètic total és de 320.25 KWh. Si estimem 188W per servidor, hores en un mes 744h, el consum mensual serà de 139.87 KWh, a dos servidors. Haurem de sumar, també, el consum que genera l'oficina durant les hores de treball. Durant aquestes 604h de treball, amb la climatització de l'oficina i il·luminació, estimem uns 600Wh per persona. Per tant, tenim un total de 362,4 KWh. En total, el consum energètic del projecte es de 822.52KWh.

## Pressupost

- Salari: Amb 604h de feina. Si estimem un cost mitjà de 15€/h per la feina d'un desenvolupador, tenim que el cost salarial del projecte és de 9060€
- Ordinador: Suposant la vida d'un ordinador d'alt rendiment és de 4 anys, i el cost d'un ordinador d'aquestes característiques és d'uns 1500€, tenim un cost per l'ordinador per hora de 0.043€/h. Per tant, el cost en ordinadors pel projecte es de 25.97€
- Consum energètic: Si tenim un consum de 320.25 kWh i el preu mitjà del kWh es de 0.135€, tenim que el cost per hora és de 43.23€. Per tant, el preu total del projecte, tenint en compte les hores invertides és de 26110.92€
- Servidors: Per cada servidor tenim un cost de 2500€ mensuals, per l'entorn d'Integració tindrem un servidor, per l'entorn de Preproducció tindrem dos. Tindrem  $2500€ * 3$  servidors ens surt un cost de 7500€

Com a cost total aproximat del projecte obtenim que és de 42696.89€

## Integració dels coneixements

### Analisis

Per prendre una decisió, o bé, analitzar les opcions que tenim, primer hem de parlar dels requeriments que hem de complir. Els requeriments, degut a que aquest desenvolupament evolutiu no ve com a petició d'un client, sinó com a millora interna del producte de l'empresa, no tenim un document de requeriments com el que es sol utilitzar en aquests casos. En canvi els requeriments ens venen de les restriccions de l'arquitectura utilitzada a producte.

Els principals requeriments que ens van demanar van ser:

- Els resultats estiguin continguts dins dels àmbits on es troben els materials de manera que cap material es mostrés a un àmbit erroni.

- Els càlculs es realitza de manera diferida, i els continguts que s'han de recomanar com a fase inicial són cursos, plans de carrera i videos formatius.
- El codi sigui lo més transversal possible per tal de que en ampliacions posteriors del mòdul, l'impacte sigui el mínim.
- El mòdul, a més, s'ha d'adaptar al patró de disseny de model-vista-controlador.

Amb aquests canvis podem saber l'impacte que pot tenir el desenvolupament en una migració a totes les aplicacions, i creiem que és viable seguir endavant per tant continuem amb les següents fases i prepararem el calendari que es seguirà a partir d'ara.

## Disseny

Per a la interacció amb els índexs, i amb el motor de lucene necessitem unes classes que serveixin d'interfície amb el motor Lucene. Aquestes interfícies s'han realitzat de manera que puguin ser extensibles a altres mòduls, com els de cerca d'entitats de la plataforma. Per això necessitarem:

- Indexador
- Cercador

Utilitzarem també models abstractes representatius de les dades que es guardaran a cada index. Les dades que guardarem als índexs hauran de ser les dades que utilitzarà Lucene com a mesura pels diferents càlculs dels criteris de recomanació. Això ens afegeix la necessitat d'introduir classes adaptadores que ens converteixin les dades que tenim guardades a base de dades, a model de dades lucene, per poder carregar-les als índexs.

Necessitarem classes controladores on hi hagi programada la lògica necesaria per a la generació de les recomanacions. Aquestes classes serán les que utilitzaran les interfícies abans mencionades. Les funcionalitats que resolen aquestes classes són:

- Ha d'interactuar amb l'indexador per tal de emmagatzemar les dades necessàries als índexs.
- Ha d'interactuar amb el cercador per tal de realitzar les consultes pertinents per a cada criteri, o situació.
- Ha de generar agrupacions seguint el algoritme de DBSCAN. Degut a què cada projecte pot contenir un número diferent de grups d'usuaris, s'ha decidit que l'algoritme d'agrupació no sigui K-means, sino que sigui determinat per a la densitat d'elements a l'espai de característiques.

Necessitarem, també, les implementacions de cada criteri. Haurem de crear els següents criteris:

- Criteri que tindrà en compte el historial de l'usuari. Utilitzant l'expedient acadèmic de l'usuari s'haurà de trobar material similar al que ja hagi realitzat.
- Criteri que tindrà en compte el grup d'usuaris on s'hi troba. Utilitzant els expedients acadèmics extreurem els materials dels usuaris agrupats al mateix grup, que l'usuari objectiu no hagi realitzat.
- Criteri que tindrà en compte els materials marcats com a favorit per l'usuari.
- Criteri que tindrà en compte els materials que l'usuari hagi avaluat amb el sistema de 5 estrelles.

Per finalitzar la part que conforma el motor del recomanador, necessitarem un procés batch que executi tot el procés per les nits i guardi a base de dades tots els .

Una vegada ja tenim definida la part del motor. Necessitarem les classes controladores on hi anirà la lògica de negoci encarregada de consultar les dades generades a base de dades. Això, i el desenvolupament de frontend ja haurem enllestit el nostre projecte.

## Model

A l'hora de definir el model de dades, haurem de distingir el model que s'utilitzarà a la base de dades Oracle, del model de dades que s'utilitzarà per emmagatzemar la informació a Lucene. El model de dades Lucene no prové informació nova, sino que utilitza informació ja existent a base de dades que serà utilitzada de manera que ens aporti informació per a les recomanacions.

Pel que fa els models de dades Lucene, i per tant, informació que ja existeix al sistema, tenim:

Usuari:

- Identificador: Utilitzarem aquest camp, per a que al realitzar les consultes mantinguem la noció de qui estem parlant, i poder relacionar-lo amb el model d'Oracle.
- Ocupació: Utilitzarem aquest camp com a camp classificador de la persona, a l'hora de generar les agrupacions d'usuaris.
- Sexe: Utilitzarem aquest camp pel mateix propòsit que el d'Ocupació. Com a camp classificador.

- Perfil d'usuari: Utilitzarem aquest camp pel mateix propòsit que el d'Ocupació. Com a camp classificador.
- Llengua: Utilitzarem aquest camp pel mateix propòsit que el d'Ocupació. Com a camp classificador.
- Localització: Utilitzarem aquest camp pel mateix propòsit que el d'Ocupació. Com a informació classificadora.

#### Cursos:

- Identificador: Utilitzarem aquest camp per a relacionar-lo amb l'objecte dins del model d'Oracle. Degut a que tots els valors que pot prendre aquest camp són diferents, no tindrà cap tipus d'impacte sobre la cerca.
- Metodologia: Aquest camp indica com s'imparteix la formació del curs, i utilitzarem el camp com a camp classificador.
- Keywords: Aquest camp l'utilitzarem com a camp classificador, la diferència amb la resta de camps classificadors que hem utilitzat és que aquest és multi-avaluat.
- Usuaris que hagin realitzat el curs
- Usuaris que hagin marcat com a favorit el curs
- Usuaris que hagin avaluat el curs amb el sistema d'estrelles

#### Plans de carrera:

- Identificador: Utilitzarem aquest camp per a relacionar-lo amb l'objecte dins del model d'Oracle. Degut a que tots els valors que pot prendre aquest camp són diferents, no tindrà cap tipus d'impacte sobre la cerca.
- Keywords: Aquest camp l'utilitzarem com a camp classificador, la diferència amb la resta de camps classificadors que hem utilitzat és que aquest és multi-avaluat.
- Cursos:
- Usuaris que hagin realitzat el curs
- Usuaris que hagin marcat com a favorit el curs
- Usuaris que hagin avaluat el curs amb el sistema d'estrelles

#### Videos formatius:

- Identificador: Utilitzarem aquest camp per a relacionar-lo amb l'objecte dins del model d'Oracle. Degut a que tots els valors que pot prendre aquest camp són diferents, no tindrà cap tipus d'impacte sobre la cerca.

- Keywords: Aquest camp l'utilitzarem com a camp classificador, la diferència amb la resta de camps classificadors que hem utilitzat és que aquest és multi-avaluat.
- Usuaris que hagin realitzat el curs
- Usuaris que hagin marcat com a favorit el curs
- Usuaris que hagin avaluat el curs amb el sistema d'estrelles

Cada model dels que hem introduït, haurà d'existir en el seu propi índex al seu sistema de fitxers diferent dels que ja existeixen. Aixó, una vegada més, s'ha de fer per a mantenir la separació del mòdul amb la resta de l'aplicatiu, i minimitzar els riscos de manera que sigui viable el projecte.

Ara definirem el model de dades que utilitzarem a la base de dades d'Oracle.

- User: Aquesta taula serà utilitzada per la majoria d'elements del modul.
- SetRecommendations: Aquesta taula es crearà i servirà com a representació d'una llista de recomanacions generada per a un usuari. Cada usuari pot tenir múltiples instàncies de SetRecommendations.
- Recommendation: Aquesta taula es crearà i s'utilitzarà com a representació d'una recomanació, on es relacionarà un SetRecommendations amb un Item, a més aquesta relació serà evaluada segons la ponderació configurada pels criteris de recomanació. També hi guardarem la relació del criteri de recomanació que ha tingut major impacte a la puntuació.
- Criteria: Aquesta taula es crearà i servirà com a representació de les diferents implementacions de Criteris. A aquesta taula s'hi incorpora el valor de ponderació sota el que es veurà afectada la puntuació retornada pel motor Lucene.

## Criteris

Els criteris que hem definit anteriorment, els explicarem amb més profunditat en aquest apartat. I parlarem concretament dels criteris que s'han desenvolupat.

### Aleatori

Degut a podem córrer el risc de que, degut a la manca d'activitat per exemple, no es generi cap recomanació per a un usuari, s'ha decidit crear un criteri base, que generarà recomanacions de manera aleatòria de continguts de la plataforma. Aquestes recomanacions estaran ponderades per un valor mínim, de manera que en el cas de que qualsevol altre criteri generi recomanacions, les generades de manera aleatòria es vegin sobreescrites. El criteri segueix el següent mètode:

- Es realitza una consulta als índexs dels materials acadèmics, demanant tots els documents inserits. Això ens retornarà tots els materials acadèmics que s'hagin guardat, que seran tots els de la plataforma.
- A cada recomanació se li otorga un valor de puntuació aleatori.
- Es retorna la llista de recomanacions.

### Criteri segons l'Expedient Acadèmic de l'Usuari

La definició d'aquest criteri es troba a l'apartat de Disseny, de manera que seguim directament amb l'explicació del mètode:

- Es realitza una consulta als índexs dels materials acadèmics, demanant tots els documents amb el identificador de l'usuari, el qual estem generant les recomanacions, al camp d'usuaris.
- Per cada document, realitzarem una cerca per als documents similars al que estem processant. Tots aquells documents similars que no continguin l'identificador de l'usuari al camp d'usuaris, el guardem per a retornar-lo com a recomanació. Com amb aquest mètode podem repetir documents, el que farem és sumar les puntuacions de totes les repeticions d'un document. Això provocarà que els materials que un usuari no hagi realitzat, però que siguin més similars a diferents materials que l'usuari si que hagi fet tingui més possibilitats de que l'usuari objectiu realitzi.
- Retornem la llista de recomanacions

### Criteri segons Similitud d'Usuaris

La definició d'aquest criteri es troba a l'apartat de [Disseny](#), de manera que seguim directament amb l'explicació del mètode:

- Es realitza una consulta de l'usuari objectiu al index d'usuaris.
- Es converteix el document de l'usuari extret de Lucene a consulta. Això s'assoleix inserint tots els atributs del document en una OR. D'aquesta manera, els documents més similars, compartiran més camps en comú, i per tant, la puntuació d'aquest document serà major.
- Es realitza una consulta de l'índex dels usuaris similars a l'usuari objectiu.
- Es crea una altra consulta, però aquest cop dels índexs dels materials acadèmics, on apareguin els identificadors dels usuaris al camp d'usuari, però que, explícitament, no aparegui l'usuari objectiu. El resultat d'aquesta consulta són els materials acadèmics que els usuaris similars a ell han realitzat i que es recomanarn a l'usuari l'objectiu, ja que ell no els ha realitzat.



- Retornem la llista de recomanacions.

### Criteri segons l'Usuari marca com a Favorit

La definició d'aquest criteri es troba a l'apartat de [Disseny](#), de manera que seguim directament amb l'explicació del mètode:

- Es realitza una consulta als índexs dels materials acadèmics, demanant tots els documents que continguin l'índex de l'usuari objectiu al camp de favorits.
- Per cada document que ens hagi retornat la consulta, realitzem una altra demanant els documents similars a aquest. Una altra vegada podem trobar el mateix material a diferents consultes al bucle. La estratègia aplicada és la mateixa que al [Criteri segons l'Expedient Acadèmic de l'Usuari](#).
- Retornem la llista de recomanacions.

### Criteri segon valoració positiva de l'Usuari

La definició d'aquest criteri es troba a l'apartat de [Disseny](#), de manera que seguim directament amb l'explicació del mètode:

- Es realitza una consulta als índexs dels materials acadèmics demanant tots els documents que continguin l'índex de l'usuari objectiu al camp de valoracions.
- Per cada document que ens hagi retornat la consulta, realitzem una altra demanant els documents similars a aquest, però aquesta vegada tindrem en compte les repeticions de l'identificador de l'usuari per a potenciar els resultats proporcionalment al número d'aparicions. D'aquesta manera els documents similars a un que s'hagi valorat més positivament, la puntuació serà més gran. I altre cop, podem trobar el mateix material a diferents consultes al bucle. La estratègia aplicada és la mateixa que al [Criteri segons l'Expedient Acadèmic de l'Usuari](#).
- Retornem la llista de recomanacions.

## Batch

El procés batch encarregat de tots els passos que s'han de fer per tal de generar les recomanacions està dividit en diferents parts:

- Càrrega de informació als índexs: A aquesta secció del codi ens haurem d'encarregar de popular els índexs de manera que aquests índexs estiguin al dia amb la nova informació que s'hagi produït al sistema.
- Generació de recomanacions: A aquesta secció del codi haurem de generar les recomanacions que siguin necessàries pel número d'usuaris que toqui generar. Limitem el número d'usuaris als que generarem les recomanacions per a no tenir un procés bloquejat que podria potencialment trigar massa temps. També limitem el número de recomanacions que guardem per a cada usuari, ja que no es disposa de

emmagatzematge il·limitat a base de dades. A més el disseny que s'ha optat per frontend a l'hora de mostrar les recomanacions, limita el número de recomanacions totals a mostrar a un valor concret, així que guardar-ne més no tindria sentit.

## Tests

Pel que fa els Tests i assegurar el funcionament del codi, s'han creat una col·lecció de tests que forcen un comportament concret del codi. D'aquesta manera si es modifiqués o s'hagués d'arreglar qualsevol cosa, el comportament no es vegi afectat. S'han desenvolupat tests unitaris, per a tots els components involucrats al motor. I també s'han desenvolupat tests d'integració.

A l'hora de fer els tests d'estrès i comprovar el comportament de tot el sistema desenvolupat es va realitzar directament sobre la plataforma, amb les dades de l'entorn d'integració del projecte de VF. Els resultats obtinguts de les diferents execucions no són realistes, ja que és un entorn intern de proves on no hi veiem activitat de cap tipus per part dels usuaris. En aquests tests ens hem trobat diferents problemes de rendiment a l'hora de tractar usuaris, ja que aquest entorn conté 100.000 usuaris. Tot i això, hem aconseguit solucionar aquests problemes paral·litzant el procés, i processant els usuaris en subconjunts. Paral·litzant hem aconseguit que el context que utilitza la plataforma i en concret la caché de hibernate es crea i es destrueix en el context del procés, de manera que amb aquesta solució aconseguim que no hi hagi degeneració en el processat de dades.

## Sostenibilitat

### Matriu de sostenibilitat

La matriu de sostenibilitat es presenta amb els apartats de Projecte en producció, vida útil i riscos. Per cada bloc tenim la valoració ambiental, econòmica i social, com es pot veure a la següent taula:

	<b>PPP</b>	<b>Vida útil</b>	<b>Riscos</b>
<b>Ambiental</b>	Consum de disseny	Empremta ecològica	Ambientals
<b>Econòmic</b>	Factura	Pla de viabilitat	Econòmics
<b>Social</b>	Impacte personal	Impacte social	Socials

## Análisis ambiental

### Consum de disseny

Com s'ha estimat a l'apartat de recursos energètics, el consum total del projecte és de 320.25kWh

### Empremta ecològica

Degut a que el projecte afegeix un procés més a realitzar durant la nit, això provoca que els càlculs realitzats durant aquest període impactin incrementant el consum energètic. A més del procés nocturn, existeix el sobrecost que aporta la nova infraestructura introduïda per visualitzar les recomanacions generades que també incrementa el consum energètic, per petició que es generi desde qualsevol client a les pàgines que inclouen alguna part d'aquest projecte.

### Riscos ambientals

Pel que fa riscos ambientals, hem de tenir en compte que les màquines que formen part de la infraestructura de l'empresa es troben a la mateixa empresa. Per tant, existeix el risc de que en el cas de que s'hi produeixi qualsevol tipus d'incident ens deixi fora de servei o que es perdin dades.

## Analisis economic

### Factura

Com hem calculat al apartat de Pressupost i càlcul de costos del projecte, el cost aproximat és de 30227.85€

### Pla de viabilitat

Degut a que els coneixements aplicats són molt especialitzats del sector de Data Science i aquest perfil no és un de necessitat primària per l'empresa, pot arribar a suposar un problema. Tant en el manteniment, com en la correcció de problemes pot impactar en una despesa d'hores superior que la que seria necessària en el cas de que el desenvolupador encarregat no tingui el coneixement.

### Riscos econòmics

En quant a riscos econòmics, s'ha de tenir en compte que en el mercat hi està havent una tendència molt clara cap a productes basats en Machine Learning i Data Science. Dita tendència ens beneficia, clarament ja que ens colocaria dins d'aquesta tendència i ens podria aportar nous clients interessats en virtagora.

Comparant aquests beneficis amb el que prèviament hem esmentat, s'assumeixen els riscos.

## Anàlisi social

### Impacte personal

Aquest projecte ens impacta positivament ja que ens introdueix en el món de Data Science i ens ha ofert la possibilitat de estudiar diferents tècniques utilitzades al mercat.

### Impacte social

Aquest projecte impacta directament a la formació dels treballadors que utilitzin la nostra plataforma com a plataforma de gestió del coneixement. Ofereix continguts acadèmics potencialment interessants per l'usuari, de manera que la probabilitat de que hi accedeixi a algun d'ells és més alta que si fos qualsevol.

### Riscos socials

S'ha de tenir en compte que si el sistema de recomanació no ofereix realment els continguts que interessin a l'usuari, pot perdre interès en el sistema, i conseqüentment pot ignorar-los. Produint que tot el sistema sigui inútil i passi a ser un destorb a la finestra de presentació de la plataforma.

### Lleis i regulacions aplicades

En qüestions legals, com es tracta amb informació d'usuaris, i la informació que generen arrel de la seva interacció amb la plataforma, estem subjectes a la llei europea del GDPR (Reglament General de Protecció de Dades). D'altra banda, val a dir que l'empresa compta amb la Certificació ISO 27001 de Sistemes de Seguretat de la Informació i, per tant, hem de seguir tots els procediments adequats a l'hora de tractar les dades. També estem subjectes als contractes de confidencialitat imposats per la mateixa empresa. Motiu pel qual molta de la informació presentada en aquesta memòria ha hagut de ser modificada o omesa.

## Conclusions

### Resultats

El projecte per part de l'empresa està en fase d'integració, de manera que encara no ha acabat aquesta fase. Aquesta fase es podrà tancar quan l'últim projecte hagi migrat a la versió on s'inclou aquest desenvolupament. S'haurà de dur a terme un seguiment i una integració esglaonada, per tal d'assegurar que no hi hagi problemes grans. A més s'ha de fer un seguiment de l'utilització d'aquest mòdul, per poder garantir l'èxit o fracàs del projecte. Les següents fases

de desenvolupament no seguiran endavant en el cas de que el mòdul no hagi sigut ben rebut, o dongui molts problemes amb les recomanacions generades. Per tal de fer un seguiment exhaustiu, s'hauria d'entregar un qüestionari als usuaris per tal de conèixer la satisfacció amb les recomanacions, i també com a feedback de la seva experiència amb el mòdul.

## Futur

Un cop tenim el mòdul desplegat a l'entorn de desenvolupament de producte, el mòdul serà introduït a la següent versió de producte, i per tant, es desplegarà amb a la resta de projectes una vegada es migrin a aquesta versió.

Un cop que succeeixi la migració només farà falta activar el mòdul i aleshores els clients tindran accés a les recomanacions, en aquell moment serà quan podrem analitzar la qualitat de les recomanacions.

Una vegada aquesta fase sigui completada, es podrà plantejar les següents fases d'ampliació del mòdul, afegint nous criteris, millorant els que ja existeixen, i afegint les opcions de configuració.

## Conclusions

Aquest projecte ens ha aportat moltes coses positives, tant com en competència comercial a nivell d'empresa, com en creixement a nivell personal. A nivell personal, val a dir que gràcies a aquest projecte he vist els meus coneixements dels meus anys a la facultat plasmat en quelcom tangible i útil.

He sigut capaç de buscar els recursos necessaris per a resoldre els requisits que se'm demanaven amb exit, gracies a les bases que m'ha proporcionat el coneixement adquirit a la carrera.

Gràcies a realitzar el treball en el context d'una empresa, amb l'ajuda d'un equip, m'ha ajudat a aprendre a treballar en equip amb tot el que això comporta. També m'ha ajudat a aprendre el valor del temps en una organització i quin és l'impacte econòmic que hi ha darrere de cada decisió. També, m'ha ajudat a veure el contrast que existeix entre el món acadèmic i el món empresarial.

## Competències tècniques aplicades

Aquest treball ha de complir amb un conjunt de competències tècniques marcades per l'especialitat de computació. Les competències tècniques amb les que el projecte compleix són les següents:

### **CCO1.3**

Definir, avaluar i seleccionar plataformes de desenvolupament i producció hardware i software per al desenvolupament d'aplicacions i serveis informàtics de diversa complexitat.

Aquesta competència s'assoleix en el procés de selecció i implementació del motor recomanador utilitzant el motor Lucene de cerca de documents. A més del desenvolupament de les interfícies necessàries per interactuar amb ell.

### **CCO2.1**

Demostrar coneixement dels fonaments, dels paradigmes i de les tècniques pròpies dels sistemes intel·ligents, i analitzar, dissenyar i construir sistemes, serveis i aplicacions informàtiques que utilitzin aquestes tècniques en qualsevol àmbit d'aplicació.

L'utilització correcta i eficient del motor Lucene de cerca requereix del coneixement previ dels paradigmes que implementa per a assolir l'objectiu de cerca massiva d'informació. L'utilització d'aquests mecanismes en favor per a generar recomanacions.

### **CCO2.4**

Demostrar coneixement i desenvolupar tècniques d'aprenentatge computacional; dissenyar i implementar aplicacions i sistemes que les utilitzin, incloent les que es dediquen a l'extracció automàtica d'informació i coneixement a partir de grans volums de dades.

El mateix sistema recomanador s'adapta a les interaccions realitzades al usuari per realitzar propostes acotades a les seves preferències. Aquesta característica del sistema compleix amb aquesta competència.

### **CCO2.5**

Implementar software de cerca d'informació (information retrieval).

Les tècniques utilitzades a la recomanació de materials, s'utilitza tècniques de cerca massiva d'informació. Com l'utilització d'indexos i el motor Lucene per a la cerca i avaluació dels materials.

### **CCO3.1**

Implementar codi crític seguint criteris de temps d'execució, eficiència i seguretat.

Les precondicions del projecte requereixen d'aquesta competència, sense l'assoliment d'aquesta competència no existiria el projecte.

## Glossari

- GEC: (Gestió del coneixement), nom de l'empresa on es va crear el projecte
- Grupo Hedima: nom de l'empresa que va absorbir a GEC
- LMS: (Learning Management System), sistema de gestió del coneixement. Software que organitza, administra, i controla les activitats de formació dels seus membres.
- Virtagora: Nom del producte que subministra l'empresa, és la part comuna de codi que implemente les funcionalitats globals i bàsiques que es comparteixen entre les aplicacions clients.
- VF: (Virtagora Factory), nom de l'aplicació client per defecte. Aquesta és l'aplicació base que s'utilitza per a desplegar i desenvolupar noves aplicacions clients.
- INT: (Integració), entorn utilitzat pels desenvolupadors pels desenvolupaments de les aplicacions.
- PRE: (Preproducció), entorn que simula l'entorn de producció on els clients hi poden realitzar les seves proves.
- PRO: (Producció), entorn real on els usuaris, i client final, tenen accés.
- Lucene: Motor de cerca massiva de documents.

## Bibliografia

Virtagora. (2019). Virtagora. [online] Available at: <http://www.virtagora.com/> [Accessed 25 Feb. 2019].

En.wikipedia.org. (2019). Recommender System. [online] Available at: [https://en.wikipedia.org/wiki/Recommender\\_system](https://en.wikipedia.org/wiki/Recommender_system) [Accessed 25 Feb. 2019].

UC San Diego, cse.ucsd.edu (2019). Amazon.com Recommendations [online] Available at: <http://cseweb.ucsd.edu/classes/fa17/cse291-b/reading/Amazon-Recommendations.pdf>

En.wikipedia.org. (2019). Collaborative filtering. [online] Available at:  
[https://en.wikipedia.org/wiki/Collaborative\\_filtering](https://en.wikipedia.org/wiki/Collaborative_filtering)

SA, G., Administradores, D., JAVIER, G., Ejecutivos, D., R..., D..., F... and B...  
(2019). Gestion Del Conocimiento Sa. [online] EINFORMA. Información de  
empresas | Informes de empresas | Datos de empresas y registros empresas.  
Available at:  
<https://www.einforma.com/informacion-empresa/gestion-conocimiento> [Accessed  
1 Jun. 2019].

expansion.com, S. (2019). Suma Capital vende el 60% de GEC al grupo Hedima.  
[online] Expansion.com. Available at:  
<http://www.expansion.com/2012/07/16/catalunya/1342465792.html> [Accessed 1  
Jun. 2019].

Catalunya, U. (2019). La UOC y "la Caixa" venden la empresa GEC a dos fondos  
de inversión. [online] Uoc.edu. Available at:  
[https://www.uoc.edu/portal/es/news/actualitat/2009/noticia\\_075.html](https://www.uoc.edu/portal/es/news/actualitat/2009/noticia_075.html) [Accessed  
1 Jun. 2019].