

CASOS D'ÚS A PROVAR: Obtenir k documents semblants

FITXERS DE PROVES NECESSARIS:

Per les proves utilitzarem el document carregats dels jocs de prova:

- *BannerFlappy.txt*
- *Maya.txt*
- *QuickAndFresh.txt*
- *SpanningTreeq.txt*
- *teoProp.txt*
- *teoTren.txt*
- *teoUni.txt*
- *Tresors1.txt*
- *Tresors2.txt*
- *VeggieFarm.txt*
- *Viajar.txt*
- *Teo.txt*

POSSIBLES CURSOS D'AQUEST CAS D'ÚS:

- Curs esperat: Seleccionen un dels documents i llavors llistem per aquell document els k documents més similars amb qualsevol de les dues opcions idf o boolean, llistem i ens mostra els k documents més rellevants.
- Cas extrem: Nombre de resultats major que el nombre de documents del sistema: Llistar i posar un nombre de documents major que el nombre de documents que té el sistema.
- Curs alternatiu: No seleccionar cap estratègia: Quan entrem a llistar no seleccionar cap estratègia i llista sense seleccionar cap de les dues.
- Curs alternatiu: Cancel·la llistar per documents similars: Quan entrem a llistar per documents similars en qualsevol moment pressionem el botó Tornar i ens hauria de tornar a la main view mostrant tots els documents sense filtratge.

- Curs alternatiu: Idiomes diferents: Quan intentem comparar dos documents, si aquests tenen idiomes diferents considerem que la seva semblança és zero, fent que ho comprovarem comparant tres documents idèntics en contingut.

ELS CASOS D'ÚS EN EL NOSTRE SISTEMA

Abans de detallar cada curs a estudiar, esmentarem els casos d'ús que proporciona: Obtenir k documents semblants. Aquest es fa des del diàleg de Llistar documents similars, on permet seleccionar el nombre de documents k, que volem que retorni i l'estratègia que ha de seguir d'on podem seleccionar també el botó Llistar per tornar a la main view i veure els k documents. Per aquest motiu, tots els cursos estudiats parteixen des del diàleg que es mostra quan seguim la ruta a l'aplicatiu: Seleccionar document-> "Llistar similars"

Gestió de documents		
Menú	Llistar per	?
Reset		
Modificar	Llistar similars	Esborrar
	Exportar	
	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo torna a la universitat	Violeta Denou
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit

Seleccionem un document

Llistar documents similars

Títol:

Tresors en un mapa (2)

Autor:

Jutge.org

Nombre de resultats:

1

Estratègia:

☐ tf-idf
☐ tf-boolean

Tornar

Llistar

Diàleg de llistar documents similars

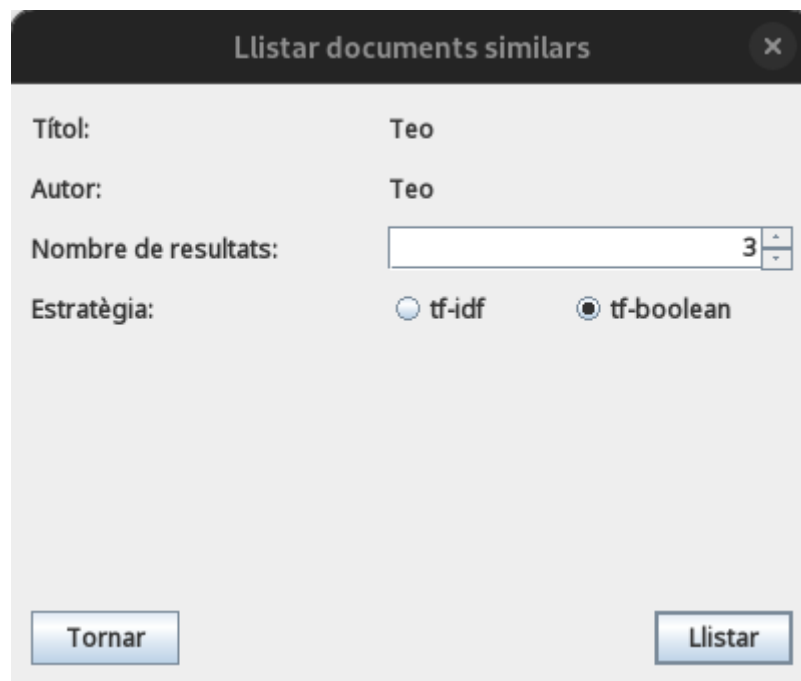
DESCRIPCIÓ DE LA PROVA:

- Curs esperat:

Per comprovar aquest curs utilitzarem el sistema inicialitzat amb tots els documents que hem comentat a l'inici menys en "Teo va a la universitat", en espanyol. Primerament, provarem l'estratègia tf-boolean pel document "Teo" amb una $k=3$.

Gestió de documents		
Menú	Llistar per	?
Reset		
<div> <div>Modificar</div> <div>Llistar similars</div> <div>Esborrar</div> <div>Exportar</div> </div>		
	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit
♥	Teo	Teo

Seleccionem un document



Llistar documents similars

Títol: Teo

Autor: Teo

Nombre de resultats: 3

Estratègia: ☐ tf-idf ☒ tf-boolean

Tornar Llistar

Inserim les dades de llistar



Gestió de documents

Menú Llistar per ? Reset

Modificar Llistar similars Esborrar Exportar

	Títol	Autor
♥	En Teo fa PROP	Ari
♥	En Teo va en tren	Violeta Denou
♥	Minimum Spanning Tree	Jordi Petit

Resultat de llistar

Per analitzar el resultat primerament comentarem que Teo només conté al contingut "Teo", en el cas d'"En Teo fa PROP" la paraula Teo apareix 1 vegada al títol + 3 vegades al contingut, en el cas d'"En Teo va en tren" la paraula Teo apareix 1 vegada al títol + 5 vegades al contingut i "Minimum Spanning Tree" on la paraula Teo no apareix.

Com que estem utilitzant ara l'estratègia tf-boolean, "En Teo fa PROP" i "En Teo va en tren" tenen el mateix pes pel fet que en els dos casos apareix la paraula Teo i l'ordre és arbitrari

seguint un paràmetre intern d'ordre en l'aplicació. Per una altra banda, "Minimum Spanning Tree" queda per avall en el resultat, ja que en ell no apareix la paraula "Teo".

Ara farem el mateix procés per l'estratègia tf-idf, per veure com canvia el resultat quan considerem la mida del text i el nombre d'aparicions.

Gestió de documents		
Menú	Llistar per ?	Reset
Modificar	Llistar similars	Esborrar
Exportar		
	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit
♥	Teo	Teo

Seleccionem un document

Llistar documents similars	
Títol:	Teo
Autor:	Teo
Nombre de resultats:	<input type="text" value="3"/>
Estratègia:	<input checked="" type="radio"/> tf-idf <input type="radio"/> tf-boolean
<div>Tornar</div> <div>Llistar</div>	

Inserim les dades de llistar

Gestió de documents		
Menú	Llistar per ?	Reset
Modificar	Llistar similars	Esborrar
Exportar		
	Títol	Autor
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Minimum Spanning Tree	Jordi Petit

Resultat de llistar

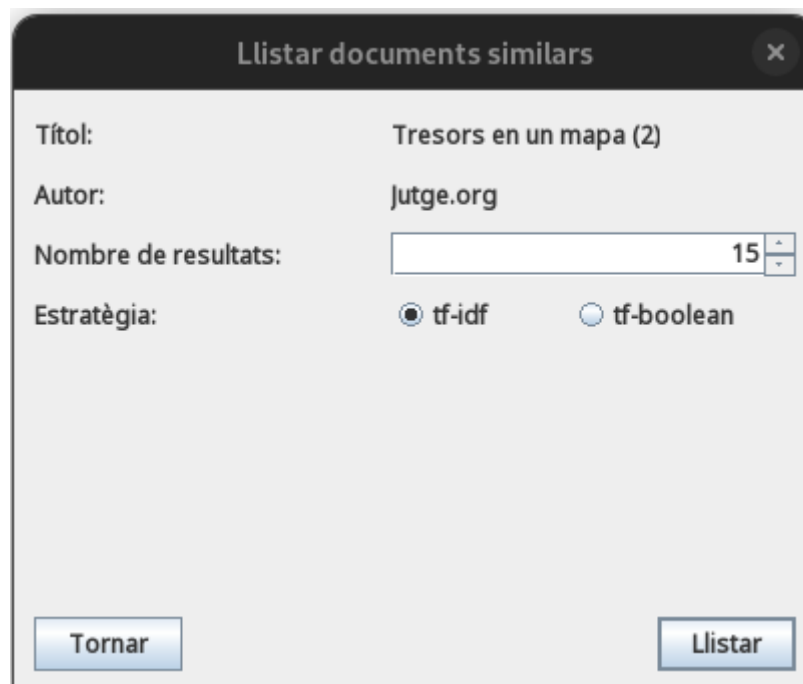
En aquest cas "En Teo fa PROP" té 50 paraules i "En Teo va en tren" 59 paraules, a pesar de l'extensió més gran del segon document, pel fet que té 6 aparicions de la paraula Teo queda per sobre del primer document, ja que té 4 aparicions. Finalment, l'últim és "Minimum Spanning Tree" perquè en ell no apareix la paraula "Teo".

- **Cas extrem: Nombre de resultats major que el nombre de documents del sistema:**

Per comprovar aquest curs utilitzarem el sistema inicialitzat amb tots els documents que hem comentat a l'inici en espanyol tots menys Minimum Spanning Tree. Entrarem al diàleg i posarem una k major que el nombre de documents carregats -1 (ja que el seleccionat no es mostra), per veure que retorna.

Gestió de documents		
Menú	Llistar per ?	Reset
Modificar	Llistar similars	Esborrar
Exportar		
	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo torna a la universitat	Violeta Denou
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit

Seleccionem un document



Llistar documents similars

Títol: Tresors en un mapa (2)

Autor: jutge.org

Nombre de resultats: 15

Estratègia: ☒ tf-idf ☐ tf-boolean

Tornar Llistar

Introduïm els paràmetres de llistar



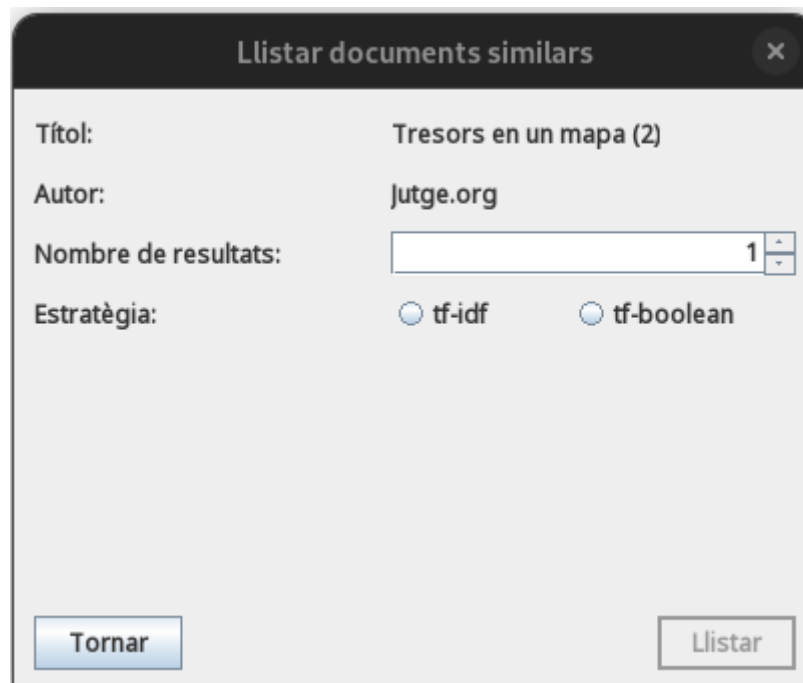
	Títol	Autor
♥	Tresors en un mapa	jutge.org
♥	En Teo torna a la universitat	Violeta Denou
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Minimum Spanning Tree	Jordi Petit
♥	Banner y Flappy	Ernest Thompson
♥	QuickAndFresh	QuickAndFresh
♥	VeggieFarm	VeggieFarm
♥	La abeja Maya	Mitsuo Kaminashi
♥	El viajero feliz	Foo Bar

Resultat de llistar

Al resultat veiem clarament com s'han mostrat tots els documents del sistema òbviament menys pel que estem llistant els seus semblants. Veiem que el resultat es troba ordenat, ja que el primer es Tresors en un mapa que té un títol semblant i el seu contingut ho és també.

- **Curs alternatiu: No seleccionar cap estratègia:**

Per comprovar aquest curs utilitzarem el sistema inicialitzat amb tots els documents que hem comentat a l'inici. Entrarem al diàleg i intentarem sense seleccionar una estratègia llistar.



Llistar documents similars

Títol: Tresors en un mapa (2)

Autor: jutge.org

Nombre de resultats: 1

Estratègia: ☐ tf-idf ☐ tf-boolean

Tornar Llistar

Inicialització del diàleg

Com veiem a la imatge d'inicialització del diàleg, el botó Llistar es troba deshabilitat fins que no hem seleccionat una de les estratègies vàlides.

- Curs alternatiu: Cancel·la llistar per autor:

Per comprovar aquest curs utilitzarem el sistema inicialitzat amb tots els documents que hem comentat a l'inici. Entrarem al diàleg i veurem si tornant ens retorna en els dos casos a la main view sense filtres.

Gestió de documents

Menú

Llistar per

?

Reset

Modificar

Llistar similars

Esborrar

Exportar

	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo torna a la universitat	Violeta Denou
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit

Estat inicial del sistema

Llistar documents similars

Títol: Tresors en un mapa (2)

Autor: jutge.org

Nombre de resultats:

Estratègia:
☐ tf-idf
☒ tf-boolean

Tornar

Llistar

Diàleg llistar documents similars

Després pressionem al botó Tornar.

Gestió de documents		
Menú	Llistar per ?	Reset
Modificar	Llistar similars	Esborrar
	Exportar	
	Títol	Autor
♥	Tresors en un mapa (2)	Jutge.org
♥	Tresors en un mapa	Jutge.org
♥	El viajero feliz	Foo Bar
♥	La abeja Maya	Mitsuo Kaminashi
♥	VeggieFarm	VeggieFarm
♥	QuickAndFresh	QuickAndFresh
♥	En Teo torna a la universitat	Violeta Denou
♥	En Teo va en tren	Violeta Denou
♥	En Teo fa PROP	Ari
♥	Banner y Flappy	Ernest Thompson
♥	Minimum Spanning Tree	Jordi Petit

Resultat de l'acció

Tornem a la main view sense haver aplicat cap filtre, que és el que esperàvem.

- **Curs alternatiu: Idiomes diferents:**

Per comprovar aquest curs utilitzarem únicament el document "El Teo fa PROP", aquest l'hem copiat tres vegades únicament canviant el títol introduint un (2) o (3), perquè no pot existir el mateix títol i mateix autor en el sistema. Inicialment, els tres documents estan al mateix idioma i compararem amb el document "En Teo fa PROP" amb $k=2$ amb estratègia tf-boolean (és indiferent una que l'altre).

Gestió de documents		
Menú	Llistar per ?	Reset
Modificar	Llistar similars	Esborrar
	Exportar	
	Títol	Autor
♥	En Teo fa PROP	Ari
♥	En Teo fa PROP (2)	Ari
♥	En Teo fa PROP (3)	Ari

Seleccionem el document

The dialog box titled "Llistar documents similars" has a close button (X) in the top right corner. It contains the following fields and controls:

- Títol:** En Teo fa PROP
- Autor:** Ari
- Nombre de resultats:** A numeric input field with the value "2" and increment/decrement buttons.
- Estratègia:** Two radio buttons: "tf-idf" (unselected) and "tf-boolean" (selected).
- At the bottom, there are two buttons: "Tornar" (left) and "Llistar" (right).

Diàleg llistar documents similars

The window titled "Gestió de documents" has standard window controls (minimize, maximize, close) in the top right. Below the title bar is a menu bar with "Menú", "Llistar per", and "?". A "Reset" button is in the top right corner. Below the menu bar are four buttons: "Modificar", "Llistar similars", "Esborrar", and "Exportar". The main area contains a table with two columns: "Títol" and "Autor".

	Títol	Autor
♥	En Teo fa PROP (3)	Ari
♥	En Teo fa PROP (2)	Ari

Resultat de l'acció

Veiem que "En Teo fa PROP (3)" queda més amunt, això és pel fet que encara tinguin la mateixa puntuació, ja que el títol i el document tenen la mateixa semblança hem creat primer "En Teo fa PROP (3)" per tant, es compara primer i queda per davant.

Ara canviem la llengua d'"En Teo fa PROP (3)" a català i tornem a fer el mateix procés inicial, per veure com en canviar l'idioma quedarà per sota d'"En Teo fa PROP (2)" perquè té diferent idioma que "En Teo fa PROP".

Modificar document [X]

Titol: En Teo fa PROP (3)

Autor: Ari

Llengua: ☒ Català ☐ Espanyol ☐ Anglès

eball en equip i altres competències transversals que també s'ha
simultaneament, i quantes hores perd solucionant bug horribles.

◀ ||| ▶

Tornar **Guardar** **Exportar**

Canviem l'idioma

Gestió de documents

Menú

Llistar per ?

Reset

Modificar

Llistar similars

Esborrar

Exportar

	Títol	Autor
	En Teo fa PROP	Ari
	En Teo fa PROP (2)	Ari
	En Teo fa PROP (3)	Ari

Seleccionem el document

The dialog box titled "Llistar documents similars" contains the following fields and controls:

- Títol:** En Teo fa PROP
- Autor:** Ari
- Nombre de resultats:** A numeric input field with the value "2" and increment/decrement buttons.
- Estratègia:** Two radio buttons: "tf-idf" (unselected) and "tf-boolean" (selected).
- Buttons:** "Tornar" (Back) and "Llistar" (List).

Diàleg llistar documents similars

The window titled "Gestió de documents" displays a table of search results. The table has two columns: "Títol" and "Autor".

	Títol	Autor
♥	En Teo fa PROP (2)	Ari
♥	En Teo fa PROP (3)	Ari

Additional UI elements include a menu bar with "Menú", "Llistar per", and "?", a "Reset" button, and a toolbar with "Modificar", "Llistar similars", "Esborrar", and "Exportar".

Resultat de l'acció

Veiem com el resultat concorda hem l'esperat, quedant ara per davant "En Teo fa PROP (2)", ja que té el mateix idioma que "En Teo fa PROP" a diferència d'"En Teo fa PROP (3)" que té un altre idioma.