

# Introducción y enunciado de la práctica análisis de chistes

16/05/2022

## Contenidos

<b>1</b>	<b>Análisis de un conjunto de chistes con metadatos</b>	<b>1</b>
1.1	Carga de datos . . . . .	1
1.2	Extracción del diccionario raw empírico desde los chistes . . . . .	3
1.3	Construcción del modelo de diccionario . . . . .	8
<b>2</b>	<b>Primer modelo de curado de los chistes</b>	<b>10</b>
2.1	Siguiente paso tratamiento de los datos curados y generación de las Document Term Matrix .	11
2.2	Generación de tópicos 4 tópicos . . . . .	12
<b>3</b>	<b>Word to vect NUEVA librería word2vec</b>	<b>14</b>
<b>4</b>	<b>Naive bayes</b>	<b>17</b>
4.1	Más chistes con metadatos . . . . .	18
<b>5</b>	<b>Enunciado</b>	<b>18</b>
5.1	Cuestión 1 . . . . .	18
5.2	Cuestión 2 . . . . .	18

## 1 Análisis de un conjunto de chistes con metadatos

Algunas ayudas y ejemplos en “Data\_model\_chistes2.Rmd”, se ha cambiado a la librería “word2vec” más reciente pero menos comentada.

El fichero “data/chistes\_con\_metadatos\_curado.csv” contiene unos 7170 chistes de la web [100chistes.com](http://100chistes.com) y de [pintamania](http://pintamania.com).

### 1.1 Carga de datos

Los datos están en un fichero separado por “;” contiene 5 variables

- origen: la web de origen del chiste; 100 chistes o pintamania **factor**
- titulo: EL título del chiste **character**.
- categoria: cortos|malos|Jaimito; son una variable **character** de categorías separadas por “|”

- palabra\_clave: políticos|argentinos; son una variable `character` de palabras clave separadas por "|" tags;
- votos: Número de votos `integer`; solo para pintamania
- texto: tipo `character`; es el texto del chiste en UTF-8 separado por " " `character`.

```
data_raw=read_csv("data/chistes_con_metadatos_curado.csv",col_names=TRUE)
```

```
## Rows: 7169 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (5): origen, titulo, categorias, palabra_clave, texto
## dbl (1): votos
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
glimpse(data_raw)
```

```
## Rows: 7,169
## Columns: 6
## $ origen      <chr> "1000 chistes", "1000 chistes", "1000 chistes", "1000 ch-
## $ titulo      <chr> "Dime con quién andas...", "Luz automática", "Política a-
## $ categorias  <chr> "cortos|malos", "cortos|malos|borrachos|matrimonios", "c-
## $ palabra_clave <chr> "feos", "neveras", "políticos|argentinos", "sangre", "fu-
## $ votos       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ texto       <chr> "- Dime con quién andas y te diré quién eres. - No ando~
```

```
knitr::kable(head(data_raw,20))
```

origen	titulo	categorias	palabra_clave	chistes	texto
1000 chistes con	Dime con quién andas...	cortos malos	feos	NA	- Dime con quién andas y te diré quién eres. - No ando con nadie... - Eres feo.
1000 chistes automáticas	Luz automática	cortos malos borrachos matrimonios	neveras	NA	Viejo marido completamente borracho y le dice a su mujer al irse para cama: - Me ha pasado algo increíble. He ido al baño y al abrir la puerta se ha encendido la luz automáticamente, sin hacer nada. - ¡La madre que te parió!, ¡Te mato!, ya te has vuelto a mear en la nevera.
1000 chistes argentinos	Política argentina	cortos malos	políticos argentinos	argentinos	Un diputado argentino se encuentra en la calle con un amigo de la infancia y éste le pregunta: - ¿Cómo estás llevando esta crisis? - ¡La verdad que duermo como un bebé! - ¡Dormís como un bebé! ¿Pero cómo hacés? - ¡Me despierto cada 3 horas llorando!
1000 chistes positivos	0	cortos malos	sangre	NA	- ¡Rápido, necesitamos sangre! - Yo soy 0 positivo. - Pues muy mal, necesitamos una mentalidad optimista.
1000 chistes porteros	Mejor portero	cortos malos	futbol porteros	NA	¿Cuál es el mejor portero del mundial? - Evidente ¡el de Para-guay!

origen	titulo	categorias	palabra_clave	texto
1000	Donación para la piscina	cortos malos	dinero agua	NA El otro día unas chicas llamaron a mi puerta y me pidieron una pequeña donación para una piscina local. Les di un garrafa de agua.
1000	Clase de astrología	cortos malos	planetas	NA - Andresito, ¿qué planeta va después de Marte? - Miércoles, señorita.
1000	Bob Esponja	cortos malos	esponja gimnasio	NA - Por qué Bob Esponja no va al gimnasio? - Porque ya está cuadrado.
1000	Ojalá lloviera	cortos malos	ciegos	NA Van dos ciegos y le dice uno al otro: - Ojalá lloviera... - Ojalá yo también...
1000	En Canarias	cortos suegras canarias	coche Noticias	NA - Noticias de última hora!! Muere una suegra atropellada en Canarias. Y esto es todo, las 8 en España y UNA menos en Canarias...
1000	Dicen que estoy loco	cortos malos	Jamón sillas	NA - Mamá, mamá, en el colegio dicen que estoy loco. - ¿Y quién dice eso de ti? - ... Me lo dicen las sillas...
1000	Bocadillo de jamón	cortos malos	madres jamón	NA - Mamá, mamá, ¿me haces un bocata de jamón? - ¿York? - Sí, túrk.
1000	Te echan de varias universidades	malos cortos	universitarias	NA - ¿Qué papis te expulsan de cuatro univerdades? - .... - Que estás perdiendo facultades
1000	Un pelo en la cama	cortos malos	cuentos pelo	NA - Qué es un pelo en una cama? - ... - El bello durmiente
1000	Entre techos	cortos malos	casas	NA - Qué le dice el techo del comedor al techo de la cocina? - .... - Te hecho de menos!
1000	Se va la luz	cortos malos	pijos escuela	NA - Qué pasa si se va la luz en una escuela privada? - .... - No se ve ni un pijo!
1000	País sin tacos	cortos malos	país	NA - En qué se convierte un país en el que se prohíben los tacos? - .... - En un país destacado!
1000	Messi de aquí a 45 días	cortos malos	deportistas Messi	NA - ¿Qué Messi en 45 días? - ..... - Mes y medio!
1000	Mundo con forma cubica	cortos malos	cubanos planeta	NA - ¿Qué pasaría si el mundo en lugar de ser una esfera fuera un cubo? - .... - Pues que todos seríamos cubanos
1000	Saludable chistes	cortos malos	amigos deportes	NA - Soy una persona muy saludable. - ¿Haces mucho deporte y comes sano? - No. Es que la gente me saluda por la calle y yo... pues les devuelvo el saludo.

## 1.2 Extracción del diccionario raw empírico desde los chistes

Extraemos al dic\_raw\_1 todas las palabras que aparecen con separación espacio.

Criterios iniciales:

- Decidimos encoding a UTF-8 columna text\_utf8 si hay que depurar por encoding habrá que ver cómo.
- Hay que decidir qué se hace con los CARACTERES ESPECIALES:{,;, () ¿?!!}. De momento los voy a eliminar
- Todas las MAYÚSCULAS a MINÚSCULAS
- De momento NO SE ELIMINAN DIGITOS: se quedan tal cual, hay que distinguir los de los dígitos de años.
- No catalogamos idiomas. ... se supone que todo está en castellano o términos técnicos que añadiremos
- Castellano es toda palabra o derivado de palabra que se encuentre en un spelling estándar de castellano que podemos ir adaptando.

```
library(tidytext)
library(stringr)
texto_df=data_raw
glimpse(texto_df)
```

```
## Rows: 7,169
## Columns: 6
## $ origen      <chr> "1000 chistes", "1000 chistes", "1000 chistes", "1000 ch-
## $ titulo      <chr> "Dime con quién andas...", "Luz automática", "Política a-
## $ categorias   <chr> "cortos|malos", "cortos|malos|borrachos|matrimonios", "c-
## $ palabra_clave <chr> "feos", "neveras", "políticos|argentinos", "sangre", "fu-
## $ votos        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ texto        <chr> "- Dime con quién andas y te diré quién eres. - No ando~
```

*#arreglo categorias a columnas distintas se podrían pasar a arrays.*

```
texto_df = texto_df %>% separate(col=c("categorias"),sep="\\|",into=paste0("C",1:5),fill="right")
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 4 rows [1015, 1039,
## 1529, 1669].
```

```
texto_df = texto_df %>% separate(col=c("palabra_clave"),sep="\\|",into=paste0("palabra",1:5),fill="right")
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 10 rows [167, 1587,
## 1589, 1657, 1988, 2072, 2190, 2233, 2363, 2376].
```

```
texto_df =texto_df %>% mutate(texto_curado=str_squish(str_replace_all(texto, "\\:|-|#|_", " ")))
glimpse(texto_df)
```

```
## Rows: 7,169
## Columns: 15
## $ origen      <chr> "1000 chistes", "1000 chistes", "1000 chistes", "1000 chi-
## $ titulo      <chr> "Dime con quién andas...", "Luz automática", "Política ar-
## $ C1          <chr> "cortos", "cortos", "cortos", "cortos", "cortos", "cortos~
## $ C2          <chr> "malos", "malos", "malos", "malos", "malos", "malos", "ma-
## $ C3          <chr> NA, "borrachos", NA, NA, NA, NA, "profesores", NA, NA, NA~
## $ C4          <chr> NA, "matrimonios", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ C5          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra1    <chr> "feos", "neveras", "políticos", "sangre", "futbol", "dine~
```

```
## $ palabra2 <chr> NA, NA, "argentinos", NA, "porteros", "agua", NA, "gimnas~
## $ palabra3 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "noticias", NA, NA, N~
## $ palabra4 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ votos <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ texto <chr> "- Dime con quién andas y te diré quién eres. - No ando ~
## $ texto_curado <chr> "Dime con quién andas y te diré quién eres. No ando con n~
```

```
## str_replace_all(text, "\\:/-/#", " ") reemplazo ":" o "-" o "#" por espacio
# esto es necesario para arreglar "hola:Pepe" que quedaría cómo una palabra si elimino:
## str_squish quita espacios repetidos
```

```
texto_tokens=texto_df %>% unnest_tokens(word, texto_curado)
glimpse(texto_tokens)
```

```
## Rows: 295,503
## Columns: 15
## $ origen <chr> "1000 chistes", "1000 chistes", "1000 chistes", "1000 chistes~
## $ titulo <chr> "Dime con quién andas...", "Dime con quién andas...", "Dime c~
## $ C1 <chr> "cortos", "cortos", "cortos", "cortos", "cortos", "cortos", "~
## $ C2 <chr> "malos", "malos", "malos", "malos", "malos", "malos", "malos"~
## $ C3 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "~
## $ C4 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "~
## $ C5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra1 <chr> "feos", "feos", "feos", "feos", "feos", "feos", "feos", "feos~
## $ palabra2 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra3 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra4 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ palabra5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ votos <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ texto <chr> "- Dime con quién andas y te diré quién eres. - No ando con ~
## $ word <chr> "dime", "con", "quién", "andas", "y", "te", "diré", "quién", ~
```

```
knitr::kable(head(texto_tokens,20))
```

origen	titulo	C1	C2	C3	C4	C5	palabra1	palabra2	palabra3	palabra4	palabra5	votos	texto	word
1000 chistes	Dime con quién andas...	cortos	malos	NA	NA	NA	feos	NA	NA	NA	NA	NA	- Dime con quién andas y te diré quién eres. - No ando con nadie... - Eres feo.	dime
1000 chistes	Dime con quién andas...	cortos	malos	NA	NA	NA	feos	NA	NA	NA	NA	NA	- Dime con quién andas y te diré quién eres. - No ando con nadie... - Eres feo.	con
1000 chistes	Dime con quién andas...	cortos	malos	NA	NA	NA	feos	NA	NA	NA	NA	NA	- Dime con quién andas y te diré quién eres. - No ando con nadie... - Eres feo.	quién

origen	título	C1	C2	C3	C4	C5	palabra1	palabra2	palabra3	palabra4	palabra5	palabra6	palabra7	palabra8	palabra9	palabra10	palabra11	palabra12	palabra13	palabra14	palabra15	palabra16	palabra17	palabra18	palabra19	palabra20	palabra21	palabra22	palabra23	palabra24	palabra25	palabra26	palabra27	palabra28	palabra29	palabra30	palabra31	palabra32	palabra33	palabra34	palabra35	palabra36	palabra37	palabra38	palabra39	palabra40	palabra41	palabra42	palabra43	palabra44	palabra45	palabra46	palabra47	palabra48	palabra49	palabra50	palabra51	palabra52	palabra53	palabra54	palabra55	palabra56	palabra57	palabra58	palabra59	palabra60	palabra61	palabra62	palabra63	palabra64	palabra65	palabra66	palabra67	palabra68	palabra69	palabra70	palabra71	palabra72	palabra73	palabra74	palabra75	palabra76	palabra77	palabra78	palabra79	palabra80	palabra81	palabra82	palabra83	palabra84	palabra85	palabra86	palabra87	palabra88	palabra89	palabra90	palabra91	palabra92	palabra93	palabra94	palabra95	palabra96	palabra97	palabra98	palabra99	palabra100	palabra101	palabra102	palabra103	palabra104	palabra105	palabra106	palabra107	palabra108	palabra109	palabra110	palabra111	palabra112	palabra113	palabra114	palabra115	palabra116	palabra117	palabra118	palabra119	palabra120	palabra121	palabra122	palabra123	palabra124	palabra125	palabra126	palabra127	palabra128	palabra129	palabra130	palabra131	palabra132	palabra133	palabra134	palabra135	palabra136	palabra137	palabra138	palabra139	palabra140	palabra141	palabra142	palabra143	palabra144	palabra145	palabra146	palabra147	palabra148	palabra149	palabra150	palabra151	palabra152	palabra153	palabra154	palabra155	palabra156	palabra157	palabra158	palabra159	palabra160	palabra161	palabra162	palabra163	palabra164	palabra165	palabra166	palabra167	palabra168	palabra169	palabra170	palabra171	palabra172	palabra173	palabra174	palabra175	palabra176	palabra177	palabra178	palabra179	palabra180	palabra181	palabra182	palabra183	palabra184	palabra185	palabra186	palabra187	palabra188	palabra189	palabra190	palabra191	palabra192	palabra193	palabra194	palabra195	palabra196	palabra197	palabra198	palabra199	palabra200	palabra201	palabra202	palabra203	palabra204	palabra205	palabra206	palabra207	palabra208	palabra209	palabra210	palabra211	palabra212	palabra213	palabra214	palabra215	palabra216	palabra217	palabra218	palabra219	palabra220	palabra221	palabra222	palabra223	palabra224	palabra225	palabra226	palabra227	palabra228	palabra229	palabra230	palabra231	palabra232	palabra233	palabra234	palabra235	palabra236	palabra237	palabra238	palabra239	palabra240	palabra241	palabra242	palabra243	palabra244	palabra245	palabra246	palabra247	palabra248	palabra249	palabra250	palabra251	palabra252	palabra253	palabra254	palabra255	palabra256	palabra257	palabra258	palabra259	palabra260	palabra261	palabra262	palabra263	palabra264	palabra265	palabra266	palabra267	palabra268	palabra269	palabra270	palabra271	palabra272	palabra273	palabra274	palabra275	palabra276	palabra277	palabra278	palabra279	palabra280	palabra281	palabra282	palabra283	palabra284	palabra285	palabra286	palabra287	palabra288	palabra289	palabra290	palabra291	palabra292	palabra293	palabra294	palabra295	palabra296	palabra297	palabra298	palabra299	palabra300	palabra301	palabra302	palabra303	palabra304	palabra305	palabra306	palabra307	palabra308	palabra309	palabra310	palabra311	palabra312	palabra313	palabra314	palabra315	palabra316	palabra317	palabra318	palabra319	palabra320	palabra321	palabra322	palabra323	palabra324	palabra325	palabra326	palabra327	palabra328	palabra329	palabra330	palabra331	palabra332	palabra333	palabra334	palabra335	palabra336	palabra337	palabra338	palabra339	palabra340	palabra341	palabra342	palabra343	palabra344	palabra345	palabra346	palabra347	palabra348	palabra349	palabra350	palabra351	palabra352	palabra353	palabra354	palabra355	palabra356	palabra357	palabra358	palabra359	palabra360	palabra361	palabra362	palabra363	palabra364	palabra365	palabra366	palabra367	palabra368	palabra369	palabra370	palabra371	palabra372	palabra373	palabra374	palabra375	palabra376	palabra377	palabra378	palabra379	palabra380	palabra381	palabra382	palabra383	palabra384	palabra385	palabra386	palabra387	palabra388	palabra389	palabra390	palabra391	palabra392	palabra393	palabra394	palabra395	palabra396	palabra397	palabra398	palabra399	palabra400	palabra401	palabra402	palabra403	palabra404	palabra405	palabra406	palabra407	palabra408	palabra409	palabra410	palabra411	palabra412	palabra413	palabra414	palabra415	palabra416	palabra417	palabra418	palabra419	palabra420	palabra421	palabra422	palabra423	palabra424	palabra425	palabra426	palabra427	palabra428	palabra429	palabra430	palabra431	palabra432	palabra433	palabra434	palabra435	palabra436	palabra437	palabra438	palabra439	palabra440	palabra441	palabra442	palabra443	palabra444	palabra445	palabra446	palabra447	palabra448	palabra449	palabra450	palabra451	palabra452	palabra453	palabra454	palabra455	palabra456	palabra457	palabra458	palabra459	palabra460	palabra461	palabra462	palabra463	palabra464	palabra465	palabra466	palabra467	palabra468	palabra469	palabra470	palabra471	palabra472	palabra473	palabra474	palabra475	palabra476	palabra477	palabra478	palabra479	palabra480	palabra481	palabra482	palabra483	palabra484	palabra485	palabra486	palabra487	palabra488	palabra489	palabra490	palabra491	palabra492	palabra493	palabra494	palabra495	palabra496	palabra497	palabra498	palabra499	palabra500	palabra501	palabra502	palabra503	palabra504	palabra505	palabra506	palabra507	palabra508	palabra509	palabra510	palabra511	palabra512	palabra513	palabra514	palabra515	palabra516	palabra517	palabra518	palabra519	palabra520	palabra521	palabra522	palabra523	palabra524	palabra525	palabra526	palabra527	palabra528	palabra529	palabra530	palabra531	palabra532	palabra533	palabra534	palabra535	palabra536	palabra537	palabra538	palabra539	palabra540	palabra541	palabra542	palabra543	palabra544	palabra545	palabra546	palabra547	palabra548	palabra549	palabra550	palabra551	palabra552	palabra553	palabra554	palabra555	palabra556	palabra557	palabra558	palabra559	palabra560	palabra561	palabra562	palabra563	palabra564	palabra565	palabra566	palabra567	palabra568	palabra569	palabra570	palabra571	palabra572	palabra573	palabra574	palabra575	palabra576	palabra577	palabra578	palabra579	palabra580	palabra581	palabra582	palabra583	palabra584	palabra585	palabra586	palabra587	palabra588	palabra589	palabra590	palabra591	palabra592	palabra593	palabra594	palabra595	palabra596	palabra597	palabra598	palabra599	palabra600	palabra601	palabra602	palabra603	palabra604	palabra605	palabra606	palabra607	palabra608	palabra609	palabra610	palabra611	palabra612	palabra613	palabra614	palabra615	palabra616	palabra617	palabra618	palabra619	palabra620	palabra621	palabra622	palabra623	palabra624	palabra625	palabra626	palabra627	palabra628	palabra629	palabra630	palabra631	palabra632	palabra633	palabra634	palabra635	palabra636	palabra637	palabra638	palabra639	palabra640	palabra641	palabra642	palabra643	palabra644	palabra645	palabra646	palabra647	palabra648	palabra649	palabra650	palabra651	palabra652	palabra653	palabra654	palabra655	palabra656	palabra657	palabra658	palabra659	palabra660	palabra661	palabra662	palabra663	palabra664	palabra665	palabra666	palabra667	palabra668	palabra669	palabra670	palabra671	palabra672	palabra673	palabra674	palabra675	palabra676	palabra677	palabra678	palabra679	palabra680	palabra681	palabra682	palabra683	palabra684	palabra685	palabra686	palabra687	palabra688	palabra689	palabra690	palabra691	palabra692	palabra693	palabra694	palabra695	palabra696	palabra697	palabra698	palabra699	palabra700	palabra701	palabra702	palabra703	palabra704	palabra705	palabra706	palabra707	palabra708	palabra709	palabra710	palabra711	palabra712	palabra713	palabra714	palabra715	palabra716	palabra717	palabra718	palabra719	palabra720	palabra721	palabra722	palabra723	palabra724	palabra725	palabra726	palabra727	palabra728	palabra729	palabra730	palabra731	palabra732	palabra733	palabra734	palabra735	palabra736	palabra737	palabra738	palabra739	palabra740	palabra741	palabra742	palabra743	palabra744	palabra745	palabra746	palabra747	palabra748	palabra749	palabra750	palabra751	palabra752	palabra753	palabra754	palabra755	palabra756	palabra757	palabra758	palabra759	palabra760	palabra761	palabra762	palabra763	palabra764	palabra765	palabra766	palabra767	palabra768	palabra769	palabra770	palabra771	palabra772	palabra773	palabra774	palabra775	palabra776	palabra777	palabra778	palabra779	palabra780	palabra781	palabra782	palabra783	palabra784	palabra785	palabra786	palabra787	palabra788	palabra789	palabra790	palabra791	palabra792	palabra793	palabra794	palabra795	palabra796	palabra797	palabra798	palabra799	palabra800	palabra801	palabra802	palabra803	palabra804	palabra805	palabra806	palabra807	palabra808	palabra809	palabra810	palabra811	palabra812	palabra813	palabra814	palabra815	palabra816	palabra817	palabra818	palabra819	palabra820	palabra821	palabra822	palabra823	palabra824	palabra825	palabra826	palabra827	palabra828	palabra829	palabra830	palabra831	palabra832	palabra833	palabra834	palabra835	palabra836	palabra837	palabra838	palabra839	palabra840	palabra841	palabra842	palabra843	palabra844	palabra845	palabra846	palabra847	palabra848	palabra849	palabra850	palabra851	palabra852	palabra853	palabra854	palabra855	palabra856	palabra857	palabra858	palabra859	palabra860	palabra861	palabra862	palabra863	palabra864	palabra865	palabra866	palabra867	palabra868	palabra869	palabra870	palabra871	palabra872	palabra873	palabra874	palabra875	palabra876	palabra877	palabra878	palabra879	palabra880	palabra881	palabra882	palabra883	palabra884	palabra885	palabra886	palabra887	palabra888	palabra889	palabra890	palabra891	palabra892	palabra893	palabra894	palabra895	palabra896	palabra897	palabra898	palabra899	palabra900	palabra901	palabra902	palabra903	palabra904	palabra905	palabra906	palabra907	palabra908	palabra909	palabra910	palabra911	palabra912	palabra913	palabra914	palabra915	palabra916	palabra917	palabra918	palabra919	palabra920	palabra921	palabra922	palabra923	palabra924	palabra925	palabra926	palabra927	palabra928	palabra929	palabra930	palabra931	palabra932	palabra933	palabra934	palabra935	palabra936	palabra937	palabra938	palabra939	palabra940	palabra941	palabra942	palabra943	palabra944	palabra945	palabra946	palabra947	palabra948	palabra949	palabra950	palabra951	palabra952	palabra953	palabra954	palabra955	palabra956	palabra957	palabra958	palabra959	palabra960	palabra961	palabra962	palabra963	palabra964	palabra965	palabra966	palabra967	palabra968	palabra969	palabra970	palabra971	palabra972	palabra973	palabra974	palabra975	palabra976	palabra977	palabra978	palabra979	palabra980	palabra981	palabra982	palabra983	palabra984	palabra985	palabra986	palabra987	palabra988	palabra989	palabra990	palabra991	palabra992	palabra993	palabra994	palabra995	palabra996	palabra997	palabra998	palabra999	palabra1000
--------	--------	----	----	----	----	----	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------

origen	título	C1	C2	C3	C4	C5	palabra1	palabra2	palabra3	palabra4	palabra5	palabra6	palabra7	palabra8	palabra9	palabra10	palabra11	palabra12	palabra13	palabra14	palabra15	palabra16	palabra17	palabra18	palabra19	palabra20	palabra21	palabra22	palabra23	palabra24	palabra25	palabra26	palabra27	palabra28	palabra29	palabra30	palabra31	palabra32	palabra33	palabra34	palabra35	palabra36	palabra37	palabra38	palabra39	palabra40	palabra41	palabra42	palabra43	palabra44	palabra45	palabra46	palabra47	palabra48	palabra49	palabra50	palabra51	palabra52	palabra53	palabra54	palabra55	palabra56	palabra57	palabra58	palabra59	palabra60	palabra61	palabra62	palabra63	palabra64	palabra65	palabra66	palabra67	palabra68	palabra69	palabra70	palabra71	palabra72	palabra73	palabra74	palabra75	palabra76	palabra77	palabra78	palabra79	palabra80	palabra81	palabra82	palabra83	palabra84	palabra85	palabra86	palabra87	palabra88	palabra89	palabra90	palabra91	palabra92	palabra93	palabra94	palabra95	palabra96	palabra97	palabra98	palabra99	palabra100	palabra101	palabra102	palabra103	palabra104	palabra105	palabra106	palabra107	palabra108	palabra109	palabra110	palabra111	palabra112	palabra113	palabra114	palabra115	palabra116	palabra117	palabra118	palabra119	palabra120	palabra121	palabra122	palabra123	palabra124	palabra125	palabra126	palabra127	palabra128	palabra129	palabra130	palabra131	palabra132	palabra133	palabra134	palabra135	palabra136	palabra137	palabra138	palabra139	palabra140	palabra141	palabra142	palabra143	palabra144	palabra145	palabra146	palabra147	palabra148	palabra149	palabra150	palabra151	palabra152	palabra153	palabra154	palabra155	palabra156	palabra157	palabra158	palabra159	palabra160	palabra161	palabra162	palabra163	palabra164	palabra165	palabra166	palabra167	palabra168	palabra169	palabra170	palabra171	palabra172	palabra173	palabra174	palabra175	palabra176	palabra177	palabra178	palabra179	palabra180	palabra181	palabra182	palabra183	palabra184	palabra185	palabra186	palabra187	palabra188	palabra189	palabra190	palabra191	palabra192	palabra193	palabra194	palabra195	palabra196	palabra197	palabra198	palabra199	palabra200	palabra201	palabra202	palabra203	palabra204	palabra205	palabra206	palabra207	palabra208	palabra209	palabra210	palabra211	palabra212	palabra213	palabra214	palabra215	palabra216	palabra217	palabra218	palabra219	palabra220	palabra221	palabra222	palabra223	palabra224	palabra225	palabra226	palabra227	palabra228	palabra229	palabra230	palabra231	palabra232	palabra233	palabra234	palabra235	palabra236	palabra237	palabra238	palabra239	palabra240	palabra241	palabra242	palabra243	palabra244	palabra245	palabra246	palabra247	palabra248	palabra249	palabra250	palabra251	palabra252	palabra253	palabra254	palabra255	palabra256	palabra257	palabra258	palabra259	palabra260	palabra261	palabra262	palabra263	palabra264	palabra265	palabra266	palabra267	palabra268	palabra269	palabra270	palabra271	palabra272	palabra273	palabra274	palabra275	palabra276	palabra277	palabra278	palabra279	palabra280	palabra281	palabra282	palabra283	palabra284	palabra285	palabra286	palabra287	palabra288	palabra289	palabra290	palabra291	palabra292	palabra293	palabra294	palabra295	palabra296	palabra297	palabra298	palabra299	palabra300	palabra301	palabra302	palabra303	palabra304	palabra305	palabra306	palabra307	palabra308	palabra309	palabra310	palabra311	palabra312	palabra313	palabra314	palabra315	palabra316	palabra317	palabra318	palabra319	palabra320	palabra321	palabra322	palabra323	palabra324	palabra325	palabra326	palabra327	palabra328	palabra329	palabra330	palabra331	palabra332	palabra333	palabra334	palabra335	palabra336	palabra337	palabra338	palabra339	palabra340	palabra341	palabra342	palabra343	palabra344	palabra345	palabra346	palabra347	palabra348	palabra349	palabra350	palabra351	palabra352	palabra353	palabra354	palabra355	palabra356	palabra357	palabra358	palabra359	palabra360	palabra361	palabra362	palabra363	palabra364	palabra365	palabra366	palabra367	palabra368	palabra369	palabra370	palabra371	palabra372	palabra373	palabra374	palabra375	palabra376	palabra377	palabra378	palabra379	palabra380	palabra381	palabra382	palabra383	palabra384	palabra385	palabra386	palabra387	palabra388	palabra389	palabra390	palabra391	palabra392	palabra393	palabra394	palabra395	palabra396	palabra397	palabra398	palabra399	palabra400	palabra401	palabra402	palabra403	palabra404	palabra405	palabra406	palabra407	palabra408	palabra409	palabra410	palabra411	palabra412	palabra413	palabra414	palabra415	palabra416	palabra417	palabra418	palabra419	palabra420	palabra421	palabra422	palabra423	palabra424	palabra425	palabra426	palabra427	palabra428	palabra429	palabra430	palabra431	palabra432	palabra433	palabra434	palabra435	palabra436	palabra437	palabra438	palabra439	palabra440	palabra441	palabra442	palabra443	palabra444	palabra445	palabra446	palabra447	palabra448	palabra449	palabra450	palabra451	palabra452	palabra453	palabra454	palabra455	palabra456	palabra457	palabra458	palabra459	palabra460	palabra461	palabra462	palabra463	palabra464	palabra465	palabra466	palabra467	palabra468	palabra469	palabra470	palabra471	palabra472	palabra473	palabra474	palabra475	palabra476	palabra477	palabra478	palabra479	palabra480	palabra481	palabra482	palabra483	palabra484	palabra485	palabra486	palabra487	palabra488	palabra489	palabra490	palabra491	palabra492	palabra493	palabra494	palabra495	palabra496	palabra497	palabra498	palabra499	palabra500	palabra501	palabra502	palabra503	palabra504	palabra505	palabra506	palabra507	palabra508	palabra509	palabra510	palabra511	palabra512	palabra513	palabra514	palabra515	palabra516	palabra517	palabra518	palabra519	palabra520	palabra521	palabra522	palabra523	palabra524	palabra525	palabra526	palabra527	palabra528	palabra529	palabra530	palabra531	palabra532	palabra533	palabra534	palabra535	palabra536	palabra537	palabra538	palabra539	palabra540	palabra541	palabra542	palabra543	palabra544	palabra545	palabra546	palabra547	palabra548	palabra549	palabra550	palabra551	palabra552	palabra553	palabra554	palabra555	palabra556	palabra557	palabra558	palabra559	palabra560	palabra561	palabra562	palabra563	palabra564	palabra565	palabra566	palabra567	palabra568	palabra569	palabra570	palabra571	palabra572	palabra573	palabra574	palabra575	palabra576	palabra577	palabra578	palabra579	palabra580	palabra581	palabra582	palabra583	palabra584	palabra585	palabra586	palabra587	palabra588	palabra589	palabra590	palabra591	palabra592	palabra593	palabra594	palabra595	palabra596	palabra597	palabra598	palabra599	palabra600	palabra601	palabra602	palabra603	palabra604	palabra605	palabra606	palabra607	palabra608	palabra609	palabra610	palabra611	palabra612	palabra613	palabra614	palabra615	palabra616	palabra617	palabra618	palabra619	palabra620	palabra621	palabra622	palabra623	palabra624	palabra625	palabra626	palabra627	palabra628	palabra629	palabra630	palabra631	palabra632	palabra633	palabra634	palabra635	palabra636	palabra637	palabra638	palabra639	palabra640	palabra641	palabra642	palabra643	palabra644	palabra645	palabra646	palabra647	palabra648	palabra649	palabra650	palabra651	palabra652	palabra653	palabra654	palabra655	palabra656	palabra657	palabra658	palabra659	palabra660	palabra661	palabra662	palabra663	palabra664	palabra665	palabra666	palabra667	palabra668	palabra669	palabra670	palabra671	palabra672	palabra673	palabra674	palabra675	palabra676	palabra677	palabra678	palabra679	palabra680	palabra681	palabra682	palabra683	palabra684	palabra685	palabra686	palabra687	palabra688	palabra689	palabra690	palabra691	palabra692	palabra693	palabra694	palabra695	palabra696	palabra697	palabra698	palabra699	palabra700	palabra701	palabra702	palabra703	palabra704	palabra705	palabra706	palabra707	palabra708	palabra709	palabra710	palabra711	palabra712	palabra713	palabra714	palabra715	palabra716	palabra717	palabra718	palabra719	palabra720	palabra721	palabra722	palabra723	palabra724	palabra725	palabra726	palabra727	palabra728	palabra729	palabra730	palabra731	palabra732	palabra733	palabra734	palabra735	palabra736	palabra737	palabra738	palabra739	palabra740	palabra741	palabra742	palabra743	palabra744	palabra745	palabra746	palabra747	palabra748	palabra749	palabra750	palabra751	palabra752	palabra753	palabra754	palabra755	palabra756	palabra757	palabra758	palabra759	palabra760	palabra761	palabra762	palabra763	palabra764	palabra765	palabra766	palabra767	palabra768	palabra769	palabra770	palabra771	palabra772	palabra773	palabra774	palabra775	palabra776	palabra777	palabra778	palabra779	palabra780	palabra781	palabra782	palabra783	palabra784	palabra785	palabra786	palabra787	palabra788	palabra789	palabra790	palabra791	palabra792	palabra793	palabra794	palabra795	palabra796	palabra797	palabra798	palabra799	palabra800	palabra801	palabra802	palabra803	palabra804	palabra805	palabra806	palabra807	palabra808	palabra809	palabra810	palabra811	palabra812	palabra813	palabra814	palabra815	palabra816	palabra817	palabra818	palabra819	palabra820	palabra821	palabra822	palabra823	palabra824	palabra825	palabra826	palabra827	palabra828	palabra829	palabra830	palabra831	palabra832	palabra833	palabra834	palabra835	palabra836	palabra837	palabra838	palabra839	palabra840	palabra841	palabra842	palabra843	palabra844	palabra845	palabra846	palabra847	palabra848	palabra849	palabra850	palabra851	palabra852	palabra853	palabra854	palabra855	palabra856	palabra857	palabra858	palabra859	palabra860	palabra861	palabra862	palabra863	palabra864	palabra865	palabra866	palabra867	palabra868	palabra869	palabra870	palabra871	palabra872	palabra873	palabra874	palabra875	palabra876	palabra877	palabra878	palabra879	palabra880	palabra881	palabra882	palabra883	palabra884	palabra885	palabra886	palabra887	palabra888	palabra889	palabra890	palabra891	palabra892	palabra893	palabra894	palabra895	palabra896	palabra897	palabra898	palabra899	palabra900	palabra901	palabra902	palabra903	palabra904	palabra905	palabra906	palabra907	palabra908	palabra909	palabra910	palabra911	palabra912	palabra913	palabra914	palabra915	palabra916	palabra917	palabra918	palabra919	palabra920	palabra921	palabra922	palabra923	palabra924	palabra925	palabra926	palabra927	palabra928	palabra929	palabra930	palabra931	palabra932	palabra933	palabra934	palabra935	palabra936	palabra937	palabra938	palabra939	palabra940	palabra941	palabra942	palabra943	palabra944	palabra945	palabra946	palabra947	palabra948	palabra949	palabra950	palabra951	palabra952	palabra953	palabra954	palabra955	palabra956	palabra957	palabra958	palabra959	palabra960	palabra961	palabra962	palabra963	palabra964	palabra965	palabra966	palabra967	palabra968	palabra969	palabra970	palabra971	palabra972	palabra973	palabra974	palabra975	palabra976	palabra977	palabra978	palabra979	palabra980	palabra981	palabra982	palabra983	palabra984	palabra985	palabra986	palabra987	palabra988	palabra989	palabra990	palabra991	palabra992	palabra993	palabra994	palabra995	palabra996	palabra997	palabra998	palabra999	palabra1000
--------	--------	----	----	----	----	----	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------

```
## [1] 23456
```

### 1.3 Construcción del modelo de diccionario

Construiremos una tabla de modelado del corpus de palabras de los chistes:

- Como primary key la word ( las `nw words`) (desde el `text_raw` en utf8)
- Su frecuencia: número de veces que aparece en los chistes
- Si es correcta según un spelling de español de España (hay que buscar... qué hay mejor)

```
count_freq=texto_tokens %>% group_by(word) %>% summarise(N=n())
dic_raw_1 = tibble(word=dic_raw_1) %>% left_join(count_freq,by="word")
```

Ahora vemos claramente cómo podemos mejorar las words para UNIFICARLAS en un único “léxico” que nos permita un tratamiento unificado, aunque las variantes escritas podrían tener significado humorístico.

Ejemplos

Palabras que contienen “zq”

```
dic_raw_1[grepl("zq",dic_raw_1$word),]
```

```
## # A tibble: 4 x 2
##   word      N
##   <chr>    <int>
## 1 izquierda  20
## 2 izquierdo   7
## 3 vazquezy    1
## 4 vezque      1
```

Palabras que contienen “ch”

```
dic_raw_1[grepl("(ch)",dic_raw_1$word),]
```

```
## # A tibble: 832 x 2
##   word      N
##   <chr>    <int>
## 1 2ºchiste    1
## 2 abolladuras.dicho 1
## 3 abrochados    1
## 4 acha        1
## 5 achedo        1
## 6 achica        1
## 7 achíiiiiiiiiis 1
## 8 achillar      1
## 9 achina        1
## 10 achiqué      1
## # ... with 822 more rows
```

Palabras (dos palabras) con :

```
dic_raw_1[grepl(":",dic_raw_1$word),]
```

```
## # A tibble: 0 x 2
## # ... with 2 variables: word <chr>, N <int>
```



### 1.3.1 Añadimos columna de spelling al diccionario

Primero veamos algunos ejemplos de las sugerencias: ver manual en de [hunspell](#). [Github diccionarios open office](#)

```
library("spelling")
library("hunspell")
#https://github.com/titoBouzout/Dictionaries # do
#es=dictionary(lang = "diccionarios/es_ES.dic", affix = "diccionarios/es_ES.dic", add_words = NULL, ca
es_ES<- dictionary("diccionarios/es_ES.dic")
#print(es_ES)
list_dictionaries()# estos son los que vienen por defecto
```

```
## [1] "bg_BG"      "ca_ES"      "cs_CZ"      "da_DK"      "de_DE_neu"  "de_DE"
## [7] "el_GR"      "en_AU"      "en_CA"      "en_GB"      "en_US"      "es_AR"
## [13] "es_BO"      "es_CL"      "es_CO"      "es_CR"      "es_CU"      "es_DO"
## [19] "es_EC"      "es_ES"      "es_GT"      "es_HN"      "es_MX"      "es_NI"
## [25] "es_PA"      "es_PE"      "es_PR"      "es_PY"      "es_SV"      "es_US"
## [31] "es_UY"      "es_VE"      "fr_FR"      "hr_HR"      "hu_HU"      "id_ID"
## [37] "it_IT"      "lt_LT"      "lv_LV"      "nb_NO"      "nl_NL"      "pl_PL"
## [43] "pt_BR"      "pt_PT"      "ro_RO"      "ru_RU"      "sh"         "sk_SK"
## [49] "sl_SI"      "sr"         "sv_SE"      "uk_UA"      "vi_VN"
```

```
hunspell_check(c("bieja","colon","colón"),dic= es_ES)
```

```
## [1] FALSE TRUE FALSE
```

```
hunspell_suggest(c("bieja","colon","colón"),dic=es_ES)
```

```
## [[1]]
## [1] "vieja" "biela"
##
## [[2]]
## [1] "colon" "clono" "colo" "colona" "colono" "colan" "colen" "color"
##
## [[3]]
## [1] "colon" "clonó" "coló" "colan" "colen"
```

```
palabras=c("amor", "amoroso", "amorosamente", "amado", "amante", "amador")
hunspell_analyze(palabras,dic=es_ES)
```

```
## [[1]]
## [1] " st:amor"      "a st:mor fl:a"
##
## [[2]]
## [1] "a st:amoroso fl:a"
##
## [[3]]
## [1] "a st:amorosamente fl:a"
##
## [[4]]
```

```
## [1] " st:amar fl:D"
##
## [[5]]
## [1] " st:amante"          " st:amantar fl:E"
##
## [[6]]
## [1] " st:amador"          "a st:mador fl:a"
```

Eliminaremos las palabras que aparezcan menos de  $K_{min} = 3$  o  $K_{max} = 500$  veces y números y tomaremos la primera sugerencia para las palabras que den incorrectas y solo la primera sugerencia.

```
K_min=3
K_max=500
dic_raw_1 = dic_raw_1 %>% filter(N>K_min & N<K_max )
dim(dic_raw_1)
```

```
## [1] 5332    2
```

```
dic_raw_1= dic_raw_1[-grep("\\w*[0-9]+\\w*\\s*",dic_raw_1$word),]
dim(dic_raw_1)
```

```
## [1] 5259    2
```

```
palabras_incorrectas= sapply(dic_raw_1$word, FUN=function(x) hunspell_check(x,dic=es_ES))
table(palabras_incorrectas)
```

```
## palabras_incorrectas
## FALSE  TRUE
## 1485  3774
```

```
lista_sugerencias= sapply(dic_raw_1$word, FUN=function(x) hunspell_suggest(x,dic=es_ES))

# nos quedamos con la primera tanto para correctas como para incorrectas

dic_raw_1$word_curada=sapply(lista_sugerencias, FUN=function(x) x[1])
dic_raw_1$lista_sugerencias=sapply(lista_sugerencias,
                                   FUN=function(x){
                                     if(length(x)>=1) {return(paste(x,collapse=","))}
                                     if(length(x)==0){return(NA)}
                                   })

# eliminamos NA

dic_raw_1 = dic_raw_1[!is.na(dic_raw_1$word_curada),]
dim(dic_raw_1)
```

```
## [1] 5238    4
```

## 2 Primer modelo de curado de los chistes

```
knitr::kable(head(dic_raw_1,20))
```

word	N	word_curadalista	sugerencias
â	5	a	a,e,o,d,u,y
aa	5	as	as,a,ara,asa,ata,ala,ama,aja,aya,ea,ar,na,ca,ta,al
aaa	10	asa	asa,ara,ata,ala,ama,aja,aya,a
aaaa	7	bezaar	bezaar
abajo	76	abajo	abajo,abajó,bajo,abaja,abaje,abano,abato,atajo,abalo,ahajo,abajá,abajé,a bajo
abanico	4	abanico	abanico,abanicó,abanicos,abanica,abanicá
abecedario	8	abecedario	abecedario,abecedarios
abeja	7	abeja	abeja,abaje,aneja,abejar,abejas,abaja,aleja
aber	12	abre	abre,saber,caber,haber,abey,ayer,aberra,rabera
abeto	4	abeto	abeto,aneto,abetos,beato,abato,abete,abito,ateto,aleto
abia	66	abiar	abiar,rabia,abina,sabia,abita,labia,abra,aria,amia,babi
abian	10	abina	abina,rabian,abinan,abitan,abiar,abran,babiano
abienta	8	avienta	avienta,ablienta,ambienta,abierta,asienta,alienta,habiente,enrabieta,tienta,entablen
abierta	8	abierta	abierta,abiertas,rabieta,abierto,acierta
abiertas	5	abiertas	abiertas,abierta,rabietas,abiertos,aciertas
abierto	5	abierto	abierto,abiertos,abierta,acierto
abiertos	4	abiertos	abiertos,abierto,abiertas,aciertos
abion	5	abino	abino,sabiondo
abitacion	6	habitaciÃ³n	habitaciÃ³n
abla	7	bala	bala,alba,ala,abala,nabla,tabla,ambla,fabla,habla,abra,arla,aula,aballa

```
texto_tokens= texto_tokens %>% right_join(dic_raw_1,word_curada,by="word")
```

## 2.1 Siguiete paso tratamiento de los datos curados y generación de las Document Term Matrix

Primera aproximación generación dela DTM del corpus de peticiones curadas. Cruzar estos datos con los tópicos/key words de losa chistes. Podéis hacerlo con tidytext o con tm (o con quanteda).

```
library(tm)
```

```
## Loading required package: NLP
```

```
##
```

```
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
## annotate
```

```
library(tidytext)
```

```
texto_tokens$N=1
```

```
DTM=cast_dtm(texto_tokens,document="titulo",term="word_curada",value=N)
```

```
MM=as.matrix(DTM)
```

```

titulos=row.names(MM)
MM=as_tibble(MM)
MM$titulo=titulos

```

## 2.2 Generación de tópicos 4 tópicos

```

library(topicmodels)
set.seed(22)
chistes_2=LDA(DTM, k=2, method = "Gibbs", control = NULL, model = NULL)

chistes_documentos <- tidy(chistes_2, matrix = "gamma")
chistes_documentos%>% arrange(document)

```

```

## # A tibble: 11,960 x 3
##   document                topic gamma
##   <chr>                  <int> <dbl>
## 1 --DAA---NI YO SE          1 0.475
## 2 --DAA---NI YO SE          2 0.525
## 3 -¿A TI QUÉ ES LO QUE MÁS TE MO  1 0.468
## 4 -¿A TI QUÉ ES LO QUE MÁS TE MO  2 0.532
## 5 -NO ME CORRIJAS          1 0.524
## 6 -NO ME CORRIJAS          2 0.476
## 7 ,METAS                   1 0.5
## 8 ,METAS                   2 0.5
## 9 !!QUE LOCO!!            1 0.483
## 10 !!QUE LOCO!!           2 0.517
## # ... with 11,950 more rows

```

```

tabla_topicos =chistes_documentos %>% pivot_wider(id_cols=document, names_from=topic,values_from= gamma,
names(tabla_topicos)[2:3]=paste0("Topico_",names(tabla_topicos)[2:3])
names(tabla_topicos)

```

```

## [1] "document" "Topico_1" "Topico_2"

```

```

Topico =apply(tabla_topicos[,2:3],1,
FUN=function(x) {
  if(x[1]>x[2]){topico=1}
  if(x[1]<x[2]){topico=2}
if(x[1]==x[2]){topico=0}
return(topico)
})

tabla_topicos = tabla_topicos %>% mutate(Clase=Topico)
tabla_topicos

```

```

## # A tibble: 5,980 x 4
##   document      Topico_1 Topico_2 Clase
##   <chr>          <dbl>   <dbl> <dbl>

```

```
## 1 Dime con quién andas...      0.561    0.439    1
## 2 Luz automática                0.416    0.584    2
## 3 Política argentina            0.463    0.537    2
## 4 0 positivo                    0.464    0.536    2
## 5 Mejor portero                 0.491    0.509    2
## 6 Donación para la piscina      0.475    0.525    2
## 7 Clase de astrología           0.519    0.481    1
## 8 Bob Esponja                   0.509    0.491    1
## 9 Ojalá lloviera                0.491    0.509    2
## 10 En Canarias                  0.467    0.533    2
## # ... with 5,970 more rows
```

Podemos extraer también las categoría o palabras clave pero son demasiadas.

```
C1=texto_df %>% select(titulo, C1)
```

```
df= C1 %>% right_join(MM,by="titulo")
names(df)[1:10]
```

```
## [1] "titulo" "C1"      "dime"    "quien"   "andas"   "eres"    "ando"    "nadie"
## [9] "feo"     "marido"
```

```
library(naivebayes)
```

```
## naivebayes 0.9.7 loaded
```

```
set.seed(1)
nrow(df)
```

```
## [1] 7133
```

```
Ntraining=floor(0.8*nrow(df))
Ntraining
```

```
## [1] 5706
```

```
Ntesting=nrow(df)-Ntraining
Ntesting
```

```
## [1] 1427
```

```
training=sample(1:nrow(df),size=Ntraining,replace = FALSE)
testing=setdiff(1:nrow(df),training)
```

```
## Warning in 1:nrow(df): numerical expression has 34074341 elements: only the first
## used
```

```
train_data=df[training,-1]
testing_data=df[testing,-c(1:2)]
```

Quizá demasiadas categorías mejor topic models a 2 , 3 o 4 ,categorías.

### 3 Word to vect NUEVA librería word2vec

<https://github.com/bnosac/word2vec>

```
#install.packages("devtools","Rtools")
#install.packages("word2vec")
```

```
library(word2vec)
txt_clean=txt_clean_word2vec(x=data_raw$texto, ascii = FALSE, alpha = TRUE, tolower = TRUE, trim = TRUE,
str(txt_clean)
```

```
## chr [1:7169] "dime con quién andas y te diré quién eres no ando con nadie eres feo" ...
```

```
model=word2vec(x=txt_clean,
  type = "skip-gram",
  dim = 50,
  window = 10,
  iter = 5L,
  lr = 0.05,
  hs = FALSE,
  negative = 5L,
  sample = 0.001,
  min_count = 5L,
  split = c(" \n,.-!?:;/\"#$%&'()*+<=>@[\\`~'{}~\t\v\f\r", ".\n?!"),
  stopwords = character(),
  threads = 1L,
  encoding = "UTF-8"
)
```

```
embedding=as.matrix(model)
emb <- predict(model, c("autobus", "jaimito", "mujer"), type = "embedding")
emb
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## autobus 0.49324778 -0.6580057 -0.9099321 -0.2372637 0.8651413 0.9308545
## jaimito 0.08688954 -2.6465790 0.3728237 -1.0025933 -0.2236835 0.8727383
## mujer 0.38372329 0.9524692 -0.2270146 -1.0718035 0.4882259 2.0631688
##           [,7]      [,8]      [,9]      [,10]      [,11]      [,12]
## autobus 1.62768722 -1.0748230 0.04890545 -0.4429723 0.003882346 0.8732639
## jaimito -0.06851102 0.1460203 -2.40517473 -0.1618383 0.116614744 -0.2503245
## mujer 0.88962364 0.4323884 -0.24232164 -1.3170614 2.190086365 -0.8013268
##           [,13]      [,14]      [,15]      [,16]      [,17]      [,18]
## autobus -1.56128728 -0.15701734 -0.6272471 -0.4105258 -0.1500595 0.06353638
## jaimito -0.08669709 -0.02753286 -0.7616298 0.5193182 2.2241023 0.38599899
## mujer -2.10406399 -0.26088908 -0.6565050 0.7424986 0.1771875 0.96865827
```

```
##           [,19]      [,20]      [,21]      [,22]      [,23]      [,24]
## autobus -2.019768 -0.0414262 -2.1455662 -1.6882231 -0.7603324 -2.2976127
## jaimito -1.641479  0.8972736 -1.8547015 -1.4017438 -0.7876659 -2.3127000
## mujer -1.633133  0.1428500 -0.8477979 -0.5069727 -0.9427141  0.4281023
##           [,25]      [,26]      [,27]      [,28]      [,29]      [,30]
## autobus -2.235957  0.745972455  1.6724601  1.0021801  1.32105947  0.6709734
## jaimito -1.676190 -0.287353635  0.8233544 -0.2391557  0.10906710  0.5161576
## mujer -1.707589  0.006950959 -0.9714662  0.4470935  0.04497885  1.1434224
##           [,31]      [,32]      [,33]      [,34]      [,35]      [,36]
## autobus -0.6470752 -0.5790763  0.1811625  0.99633789 -0.008319272  0.1551851
## jaimito -0.9901198 -0.1531859  0.5353013  0.50374371 -1.186450720  0.3026336
## mujer  0.1178002  0.4514101  0.2480725  0.09229566  0.737223089  2.4230177
##           [,37]      [,38]      [,39]      [,40]      [,41]      [,42]
## autobus -0.7236770 -0.1784567 -0.1565309 -0.004147133  0.1433230  0.26644439
## jaimito -0.9294956 -0.7300665  0.4991387  0.711062431  0.4864651 -0.07855427
## mujer -1.1902472 -0.2207187 -1.7618690  0.886426568 -0.4059232  0.47679093
##           [,43]      [,44]      [,45]      [,46]      [,47]      [,48]
## autobus -0.9399630  0.3532828  1.070434  1.2255282 -1.3430599  0.4361992
## jaimito -0.9245359 -1.2500688 -1.076714 -0.3954946 -0.2281055  0.6138368
## mujer -0.8164394 -0.2584829 -0.812281  0.5730899  0.4066224 -1.3106833
##           [,49]      [,50]
## autobus -0.6812266  1.2139244
## jaimito -0.2570671  0.8486603
## mujer  1.5039954  0.4146460
```

```
nn <- predict(model, c("jaimito", "profesor"), type = "nearest", top_n = 5)
nn
```

```
## $jaimito
##      term1      term2 similarity rank
## 1 jaimito      anota  0.8632550    1
## 2 jaimito      jamito  0.8625439    2
## 3 jaimito      traerne 0.8558403    3
## 4 jaimito      jeimito 0.8520830    4
## 5 jaimito oraciones 0.8512632    5
##
## $profesor
##      term1      term2 similarity rank
## 1 profesor      memin  0.9005938    1
## 2 profesor      alumno 0.8990849    2
## 3 profesor      examen 0.8990266    3
## 4 profesor      frutas 0.8933191    4
## 5 profesor      marte  0.8789321    5
```

```
doc2vec(model, c("padre", "madre", "hijo"))
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,]  0.2727478 -0.3490508  1.5674843 -1.678564 -0.2206307  1.4065835  0.2302239
## [2,] -0.5575424 -1.5284566  1.0829769 -1.975983 -0.1349357  1.3036857  1.7181846
## [3,] -0.2595362  0.3817862  0.5053649 -0.399771  0.5186067  0.7885997  1.3912740
##           [,8]      [,9]      [,10]      [,11]      [,12]      [,13]
## [1,]  0.2487679  0.465802 -0.05547563  0.1768579  0.03859647 -1.4896098
## [2,]  0.6846665 -0.796141  0.20620546  0.7283186 -0.44242108 -0.7485540
```

```
## [3,] -0.4407749 -1.011690 0.97146793 -1.1738757 -0.28388912 -0.7907744
##      [,14]      [,15]      [,16]      [,17]      [,18]      [,19]      [,20]
## [1,] -0.8444553 -1.691198 0.09345359 0.54973028 1.2397960 -1.250630 1.255210
## [2,] 0.3037887 -1.556308 0.75041711 0.04560138 0.3607031 -2.226147 1.542805
## [3,] -1.3868917 -1.243202 0.41086037 0.22678809 0.4006453 -1.772973 1.533564
##      [,21]      [,22]      [,23]      [,24]      [,25]      [,26]      [,27]
## [1,] 0.2897654 -2.437773 0.7739895 -2.107992 -2.0615837 -0.1313548 1.164457
## [2,] -0.4911121 -1.951823 0.3802617 -1.681261 -1.1677467 -1.0376462 2.084226
## [3,] -0.6522840 -2.700715 1.0940609 -1.119187 -0.8018186 -0.9855824 1.633107
##      [,28]      [,29]      [,30]      [,31]      [,32]      [,33]
## [1,] 1.5201086 0.6323793 -0.7624062 -0.6160427 -0.7656532 -0.6722285
## [2,] -0.4120705 0.1677049 0.9504106 -0.5228955 0.3909138 0.1997444
## [3,] 1.0284834 1.7344975 0.1017707 -1.7265351 0.2119278 0.6182938
##      [,34]      [,35]      [,36]      [,37]      [,38]      [,39]
## [1,] -0.1387076 1.0478117 0.06770922 -0.4950438 -0.3462227 -0.6745713
## [2,] -0.1948977 -1.0467991 1.40255654 -0.1413882 0.7476683 0.5385075
## [3,] 0.5490976 -0.1692591 0.57563527 -0.3437360 0.3434265 -0.0124320
##      [,40]      [,41]      [,42]      [,43]      [,44]      [,45]      [,46]
## [1,] -0.7162034 1.4135215 -0.2275291 -0.9153525 -1.116554 -0.4442853 1.1567313
## [2,] 0.6619150 0.1618558 -0.5189599 -1.1197389 -1.382463 -0.6134729 -0.6250816
## [3,] 1.0126826 1.4944657 0.1847321 -0.8298018 -1.948009 -0.4662968 0.5148879
##      [,47]      [,48]      [,49]      [,50]
## [1,] -0.72215962 0.4111029 -0.87880809 0.6041034
## [2,] -0.07010143 -0.7365789 0.09191219 -0.2510743
## [3,] 0.13146706 -0.5350725 -0.33289758 1.2036003
```

```
M=as.matrix(model)
dim(M)
```

```
## [1] 4375 50
```

```
#Simi=word2vec_similarity(M,M,top_n=+Inf, type="cosine")
cosine <- function(x,y) sum(x * y)/sqrt(sum(x^2)*sum(y^2))
# install.packages("proxy")
library(proxy)
```

```
##
```

```
## Attaching package: 'proxy'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      as.dist, dist
```

```
## The following object is masked from 'package:base':
```

```
##
```

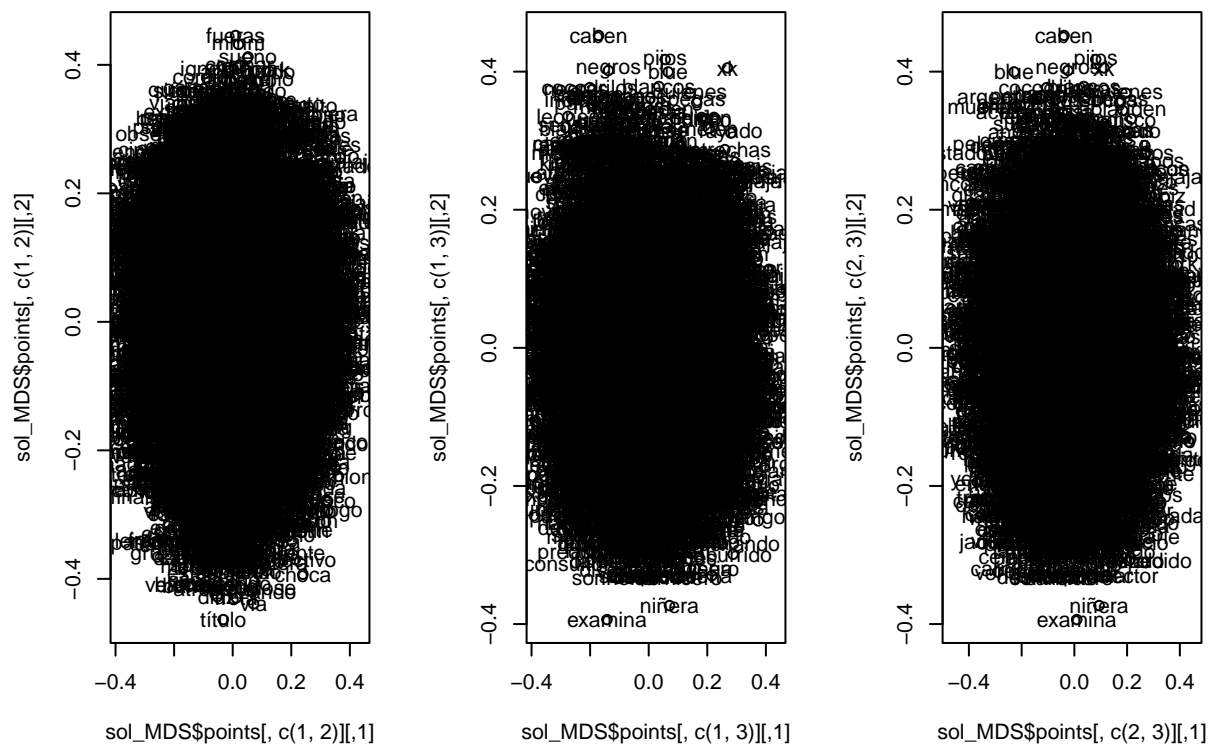
```
##      as.matrix
```

```
SS=as.matrix(simil(M,method=cosine))
diag(SS)=1
D=sqrt(1-SS)
dimnames(D)=list(dimnames(M)[[1]],dimnames(M)[[1]])
sol_MDS=cmdscale(D,k = 3,list=TRUE)
str(sol_MDS)
```



```
## List of 5
## $ points: num [1:4375, 1:3] -0.3694 -0.0763 0.1059 0.1183 0.0785 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:4375] "usd" "tocas" "ria" "caducado" ...
## .. ..$ : NULL
## $ eig : NULL
## $ x : NULL
## $ ac : num 0
## $ GOF : num [1:2] 0.243 0.243
```

```
par(mfrow=c(1,3))
plot(sol_MDS$points[,c(1,2)])
text(sol_MDS$points[,c(1,2)],dimnames(M)[[1]])
plot(sol_MDS$points[,c(1,3)])
text(sol_MDS$points[,c(1,3)],dimnames(M)[[1]])
plot(sol_MDS$points[,c(2,3)])
text(sol_MDS$points[,c(2,3)],dimnames(M)[[1]])
```



```
par(mfrow=c(1,1))
```

## 4 Naive bayes

Podéis utilizar algún algoritmo de naivebayes con los metadatos de los chistes (archivo que se explica abajo) o con topic models.

## 4.1 Más chistes con metadatos

En el fichero de este git “chistes\_con\_metadatos.csv” hay más chistes con dos columnas de metadatos para practicar.

# 5 Enunciado

Basándonos en las ayudas de Enunciado\_taller2\_chistes\_con\_metadatos.Rmd" lo anterior generar un modelo de datos con 4 tópicos (de topic models o combinado con categorías o palabras clave. Asignar cada tópico a su  $\alpha$  más alto) y un diccionario de palabras curadas por chistes.

## 5.1 Cuestión 1

Naive Bayes para predecir las 4 categorías de chistes a partir de las variables de presencia ausencia de las palabras. Evaluar el modelo.

## 5.2 Cuestión 2

A partir de la librería `word2vec` generar una proyección y estudiar si las palabras se agrupan según en las proyecciones