

Documento 1

Batalla América en el 2-1 ante Atlético San Luis La ausencia de Katty Martínez ...				
Corpus Document	Vector Representation	Extracted Features	Comparison Element	Similarity
241	Binarized	Bigram	Title	0.447
241	Frequency	Bigram	Title	0.447
241	TF-IDF	Unigram	Title	0.447
146	Binarized	Unigram	Title	0.303
226	Frequency	Unigram	Title + Content	0.302
147	Frequency	Unigram	Title + Content	0.301
226	Frequency	Unigram	Content	0.3
189	Binarized	Unigram	Content	0.28
147	Binarized	Unigram	Content	0.279
147	Binarized	Bigram	Content	0.279

Documento 2

Bryn Terfel se presentará en el Palacio de Bellas Artes				
Corpus Document	Vector Representation	Extracted Features	Comparison Element	Similarity
132	TF-IDF	Unigram	Title + Content	0.68
132	TF-IDF	Unigram	Title	0.645
132	Binarized	Unigram	Title	0.596
132	Frequency	Unigram	Title	0.596
132	Frequency	Unigram	Title + Content	0.589
132	TF-IDF	Unigram	Content	0.537
132	Binarized	Bigram	Title	0.5
132	Frequency	Bigram	Title	0.5
132	TF-IDF	Unigram	Title	0.5
132	Binarized	Unigram	Content	0.5

Documento 3

Durante el año pasado, los mexicanos invirtieron 6 mil 429 millones de dólares en ...				
Corpus Document	Vector Representation	Extracted Features	Comparison Element	Similarity
2	Frequency	Unigram	Content	0.993
2	TF-IDF	Bigram	Content	0.989
2	Binarized	Bigram	Content	0.988
2	Frequency	Bigram	Content	0.988
2	Binarized	Unigram	Content	0.987
2	TF-IDF	Unigram	Content	0.979
2	Frequency	Unigram	Title + Content	0.956
2	Binarized	Unigram	Title + Content	0.94
2	Binarized	Bigram	Title + Content	0.917
2	Frequency	Bigram	Title + Content	0.917

Documento 4

Pemex ha recuperado gran parte de ventas en el mercado interno				
Corpus Document	Vector Representation	Extracted Features	Comparison Element	Similarity
8	Binarized	Unigram	Title	1.0
8	Binarized	Bigram	Title	1.0
8	Frequency	Unigram	Title	1.0
8	Frequency	Bigram	Title	1.0
8	TF-IDF	Unigram	Title	1.0
8	TF-IDF	Unigram	Title	1.0
8	TF-IDF	Unigram	Title + Content	0.746
8	Frequency	Unigram	Title + Content	0.724
8	Frequency	Bigram	Title + Content	0.72
8	TF-IDF	Bigram	Title + Content	0.717

Documento 5

Durante el año pasado, los mexicanos invirtieron 6 mil 429 millones de dólares en ...				
Corpus Document	Vector Representation	Extracted Features	Comparison Element	Similarity
2	Frequency	Unigram	Content	0.993
2	TF-IDF	Bigram	Content	0.989
2	Binarized	Bigram	Content	0.988
2	Frequency	Bigram	Content	0.988
2	Binarized	Unigram	Content	0.987
2	TF-IDF	Unigram	Content	0.979
2	Frequency	Unigram	Title + Content	0.956
2	Binarized	Unigram	Title + Content	0.94
2	Binarized	Bigram	Title + Content	0.917
2	Frequency	Bigram	Title + Content	0.917