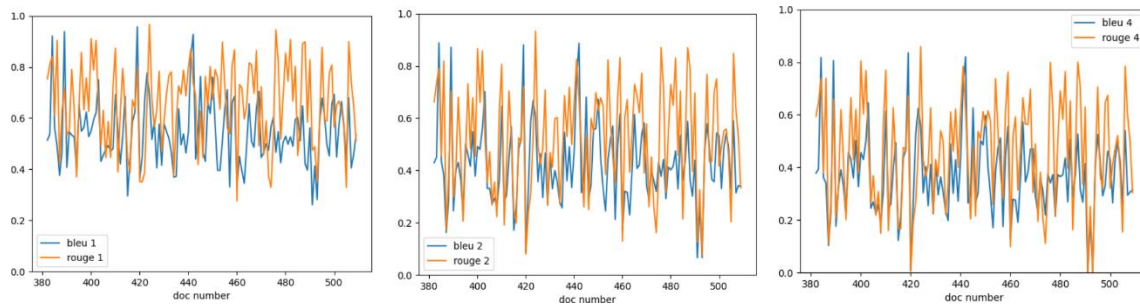# Tema 3 IA

Rosca Nicolae 343C3

Analiza cantitativa

Bayes with stop words

Acesta este bayes default, no fancy stuff no nothing.


Average bleu 0.5411218507106947 average rouge 0.6569459218116729 for k = 1

Average bleu 0.4223318419532274 average rouge 0.5226551133149004 for k = 2

Average bleu 0.36930204963377145 average rouge 0.4565964695123247 for k = 4

Bayes without stop words

Eliminarea stopwords-urilor, in mod complet neasteptat, ofera un sumar care este aproximativ la fel de concis (scorul bleu) insa tinde sa contina fraze mai relelvante (scorul rouge). De ex: pt k = 1 avem 0.69 vs 0.65 pentru rouge, o diferenta ce nu poate fi neglijat.
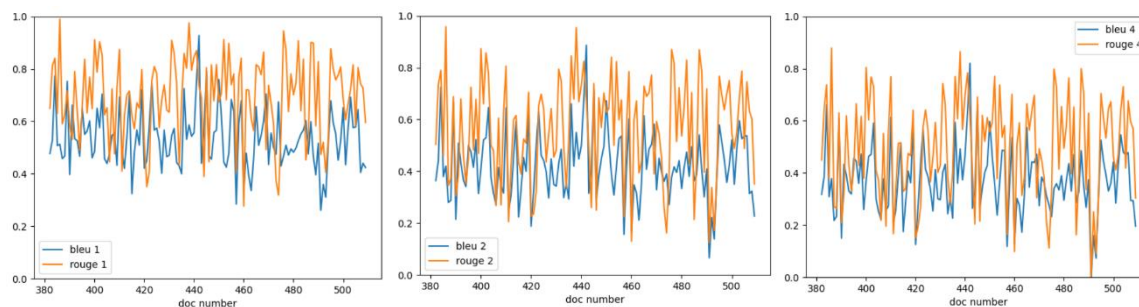
Average bleu 0.5335882543319638 average rouge 0.6928991716701964 for k = 1

Average bleu 0.4211043649113364 average rouge 0.5657101421367023 for k = 2

Average bleu 0.36877658616384223 average rouge 0.4969321671886883 for k = 4

Bayes without stop words and with lemantization

Adugarea procesului de lemantizare in pipe-line-ul spacy duce la o crestere si mai mare a relevantei continutului generat, iar natura concisa este pastrata.
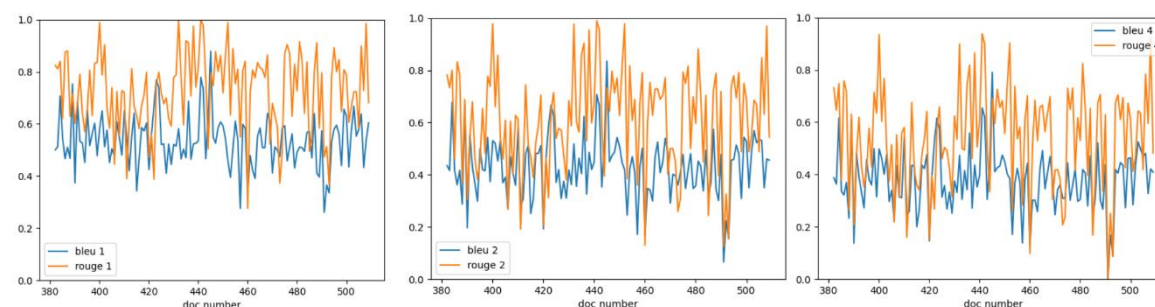

Average bleu 0.5353925316085677 average rouge 0.7275347559495742 for k = 1

Average bleu 0.4289253933338591 average rouge 0.6095843058437675 for k = 2

Average bleu 0.3776141558896723 average rouge 0.5393915433134657 for k = 4

Tf-idf with all features (noun, title similarity, weight, sent_count == 3, and similarity_weight = 0.8)

Acesta este baseline-ul. Vom folosi datele pentru a compara restul modificarilor aduse.
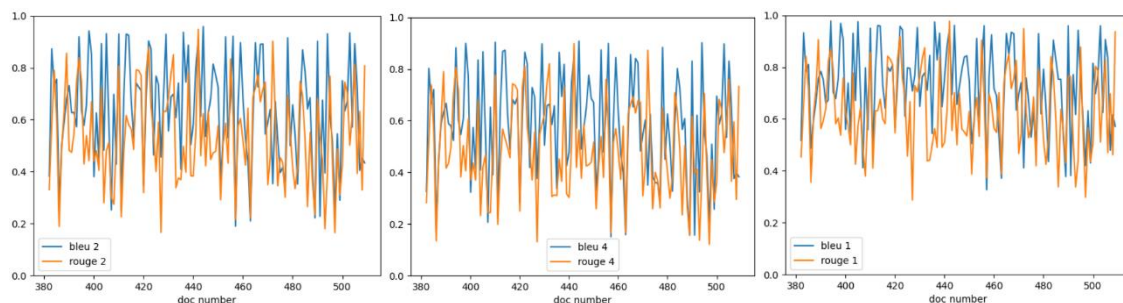
Average bleu 0.727058594270946 average rouge 0.6309704310230276 for k = 1

Average bleu 0.6347745332373093 average rouge 0.5370334640967181 for k = 2

Average bleu 0.583405981756282 average rouge 0.48808214823634805 for k = 4

Tf-idf without noun

Eliminarea procesului de selectare numai al substantivelor aduce penalizari drastice: 0.72 vs 0.65 pentru bleu si 0.63 vs 0.59.
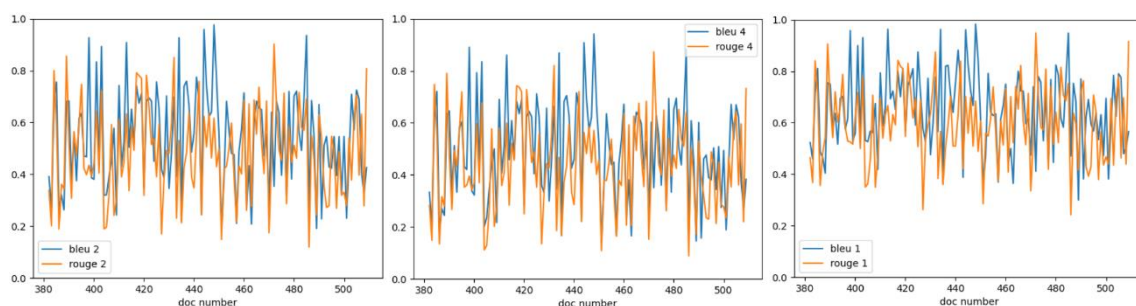

Average bleu 0.6531608203464198 average rouge 0.5990211478420028 for k = 1

Average bleu 0.5448400904860323 average rouge 0.48814588472364406 for k = 2

Average bleu 0.4978520970187487 average rouge 0.4398577552241047 for k = 4

Tf-idf without title similarity

Daca eliminam factorul de similaritate cu titlu vom obtine sumare care tin sa fie la fel de concise, insa nu au continutul relevant. Este de asteptat, deoarece in titlu tind sa existe cuvinteele cele mai relevante pentru sumar.
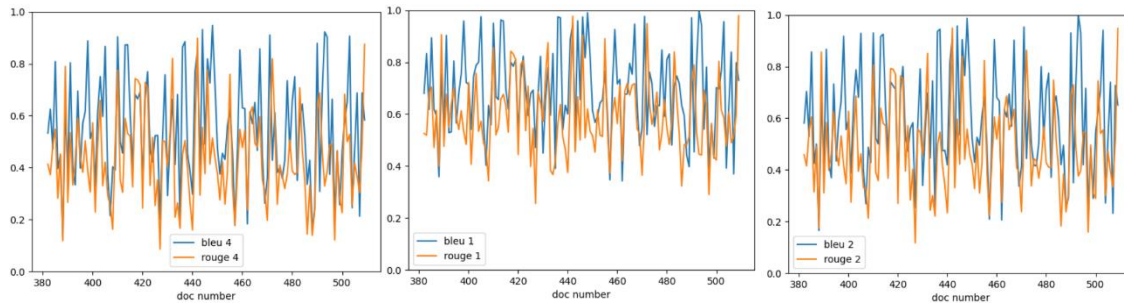
Average bleu 0.6924811383520532 average rouge 0.5910421232921976 for k = 1

Average bleu 0.594054708717342 average rouge 0.4823157216134726 for k = 2

Average bleu 0.5453463174814798 average rouge 0.434707055301333 for k = 4

Tf-idf without custom made weight

Aici am stat destul de mult sa gasesc o functie de weight ce ofera rezultate peste baseline. Am considerat ca cuvintele relevante tind sa stea la inceputul articolului (titlu ex si introducerea), prin urmare am folosit functia y = -x/6 + 1. Dupa cum se poate obserca in grafic, la x = 0 avem valoarea cea mai mare (in domeniul 0, 1), si scade relativ incet.
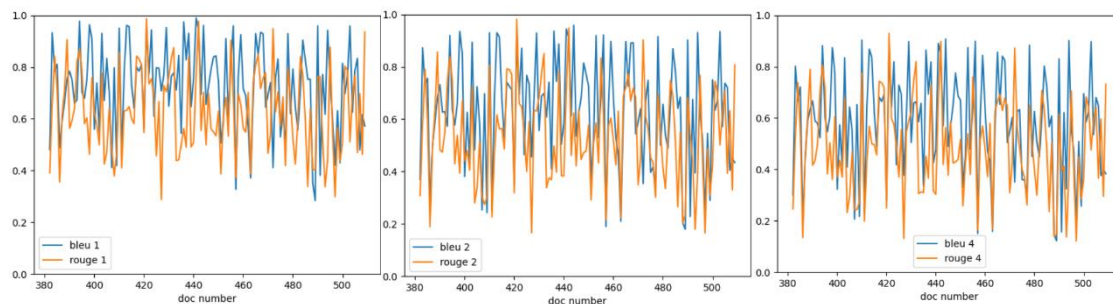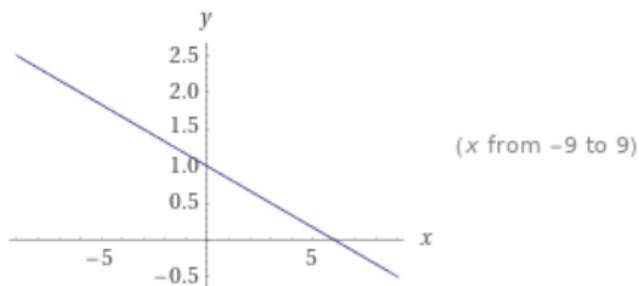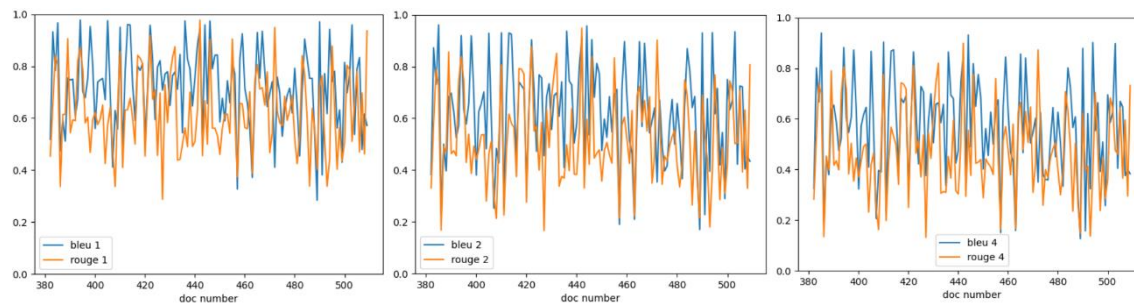

Average bleu 0.7242713433558859 average rouge 0.6280303155146706 for k = 1

Average bleu 0.6316283958817295 average rouge 0.5339087608735084 for k = 2

Average bleu 0.5806333438536918 average rouge 0.4855368322172102 for k = 4

Tf-idf with title similarity_weight == 0.4

Daca folosim in title weight prea mic avem rezultate, per total, mai rele.

Average bleu 0.7180903765580317 average rouge 0.6209359668714137 for k = 1

Average bleu 0.6219240525343053 average rouge 0.5228237324010725 for k = 2

Average bleu 0.5714652622979456 average rouge 0.4741257845635932 for k = 4

Tf-idf with title similarity_weight == 1.2

O supraaxare pe titlu, de asemenea este o idee rea. Am observat valoarea de 0.8 pentru similarity_weight ca find ideala.
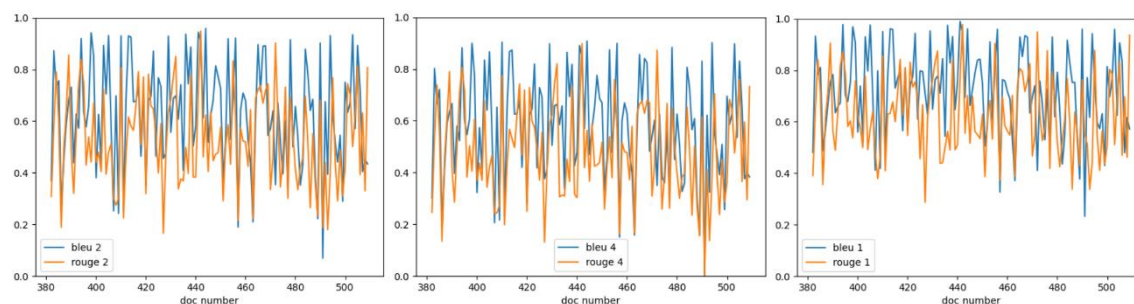
Average bleu 0.7222199303471122 average rouge 0.624784802906049 for k = 1

Average bleu 0.6294148218845569 average rouge 0.5295149370314338 for k = 2

Average bleu 0.5788449332062278 average rouge 0.48061124932773175 for k = 4

Tf-idf with title sent_count == 3

Un numar mai mic de fraze duce la un sumar mai concis, insa fara asa de multe informatii.
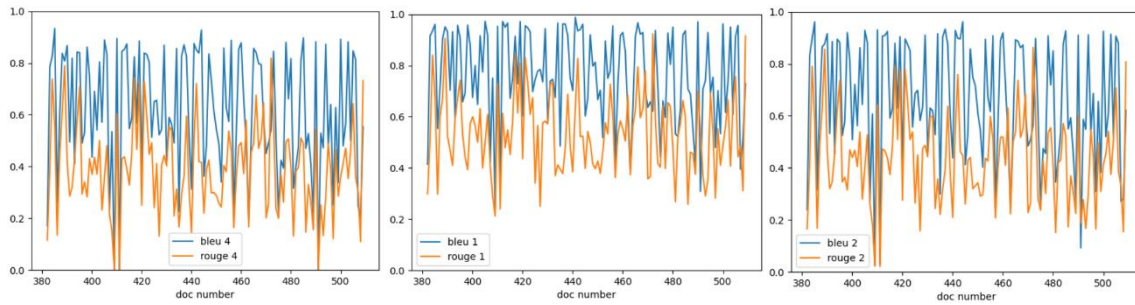
Average bleu 0.7604902645304347 average rouge 0.5351899255434628 for k = 1

Average bleu 0.6626974666047125 average rouge 0.43302363087793594 for k = 2

Average bleu 0.6131849927689007 average rouge 0.3906398236015485 for k = 4

Tf-idf with title sent_count == 5

Un sumar cu mai multe fraze duce la o cantitae mai mare de informatii relevante, insa sumarul devine mai putin concis.
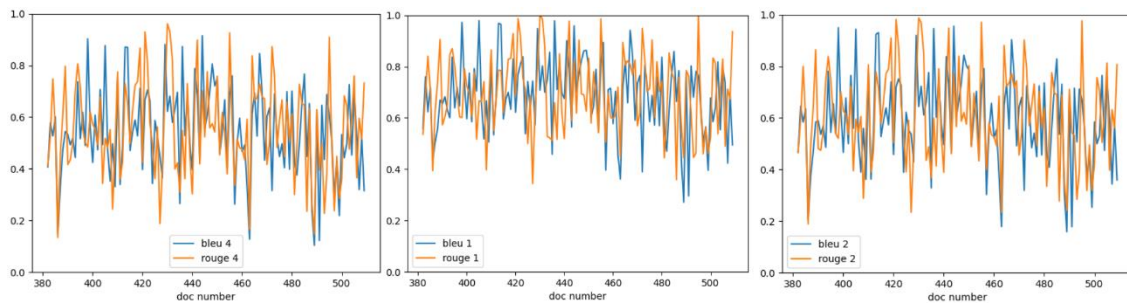
Average bleu 0.6859301023697069 average rouge 0.6953635581922839 for k = 1

Average bleu 0.5934315727504412 average rouge 0.6069237275867106 for k = 2

Average bleu 0.5436219756242754 average rouge 0.5545842032287045 for k = 4
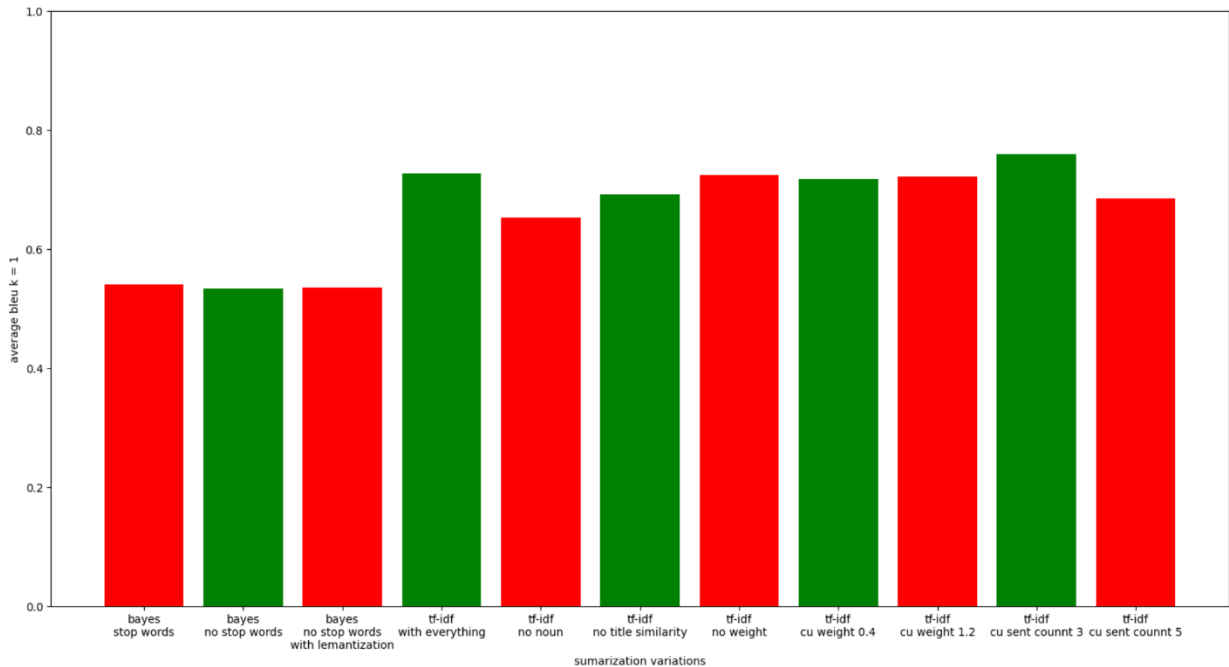
Din graficul de mai sus, reises facptul ca tf-idf tinde sa produca sumarele cele mai concise.
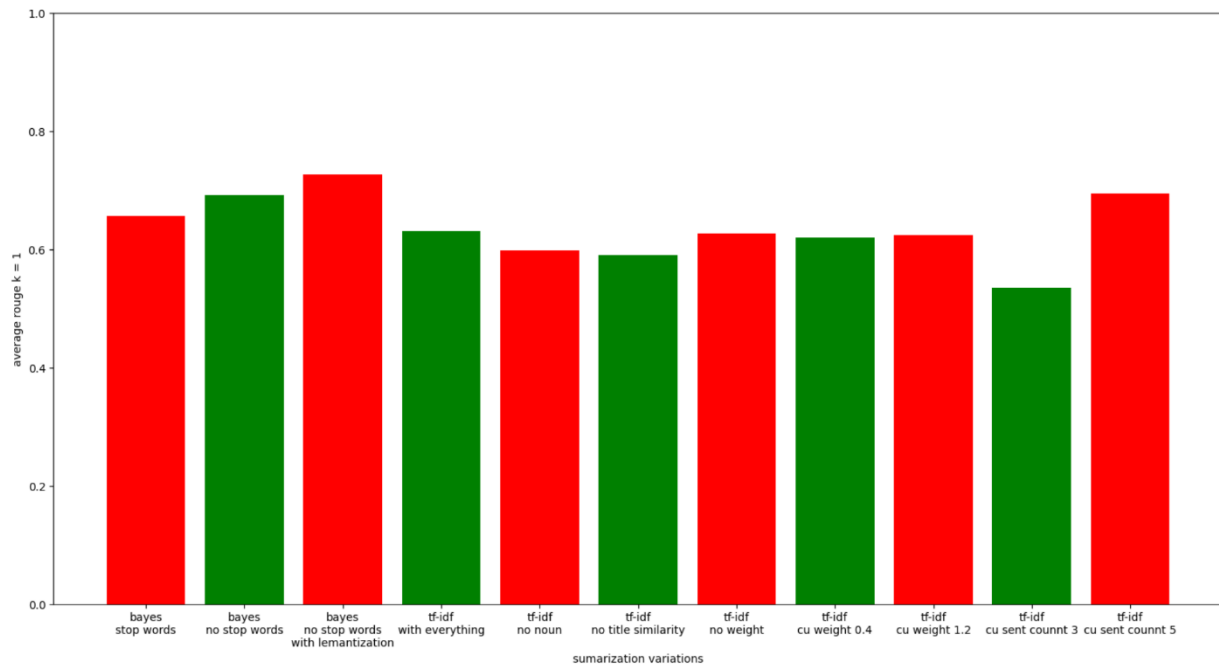
Iar in acest grafic reiese metodele care intorc cantitatea cea mai mare de informatii relevante pentru un sumar. Castigatorii sunt bayes cu taote modulele su tf-idf cu toate modulele si un sentence_count de 5!

Analiza calitativa

Pentru Bayes cu eliminare de stopwords si tokenizare. Gold standard-ul in aceasta tema pentru sumarizare automata. Eliminarea oricarui proces din pipeline duce la producerea sumarelor care sunt mai putin relevant.

Articol bussiness

Low-cost airlines hit Eurotunnel

Channel Tunnel operator Eurotunnel has seen sales fall in the face of the upsurge in European low-cost airlines.

The firm said sales were down 4% in 2004 to 789m euros ($1.03bn; £548m). "The impact of the development of no-frills airlines is being felt ever more strongly," said chief executive Jean-Louis Raymond. Income from its vehicle-carrying shuttle services fell 7%, although 15% more passengers meant a 2% rise in railway revenue.

The cross-Channel truck market is improving, Eurotunnel said, but warned that it was not benefiting since

much of the traffic was in containers destined for ports. The passenger-only trains which use the tunnel are run by a separate company, Eurostar.

Eurotunnel is still struggling with debts of more than 6bn euros. The company is currently kept afloat by

the 200-plus banks to whom it owes the money. A shareholder revolt threw out the old board in 2004. But the BBC's business editor, Jeff Randall, said the banks could yet step in and take over altogether. "At the

moment it can't even service the interest on its debt," he said. "This is a company in the departure lounge of life."

Sumar uman

The firm said sales were down 4% in 2004 to 789m euros ($1.03bn; £548m)."At the moment it can't even service the interest on its debt," he said.Eurotunnel is still struggling with debts of more than 6bn euros.But the BBC's business editor, Jeff Randall, said the banks could yet step in and take over altogether.The cross-Channel truck market is improving, Eurotunnel said, but warned that it was not benefiting since much of the traffic was in containers destined for ports.

Sumar automat bayes cu eliminare stopwords si lemantizare

Low-cost airlines hit Eurotunnel

Channel Tunnel operator Eurotunnel has seen sales fall in the face of the upsurge in European low-cost airlines.

The firm said sales were down 4% in 2004 to 789m euros ($1.03bn; £548m).

The cross-Channel truck market is improving, Eurotunnel said, but warned that it was not benefiting since

much of the traffic was in containers destined for ports.

Eurotunnel is still struggling with debts of more than 6bn euros."At the moment it can't even service the

interest on its debt," he said."This is a company in the departure lounge of life."

Articol entertainment

Lost Doors frontman movie found

Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.

The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had

been rejected by Florida State University. Morrison is seen quizzing a college administrator on why he was refused. "But what happened? How come my parents and the state and the university didn't look ahead?" he

is seen asking.

"It's incredible. He's so clean cut and soft-spoken," said Florida state archivist Jody Norman. "We know he was at Florida State University for a period of time and he did some acting when he was there," Norman added. The Doors were one of the most influential bands of the 1960s, with hits including Light My Fire and Riders On The Storm. Morrison was notorious for his wild lifestyle - and was accused of exposing himself

and simulating a sex act at a Miami concert in 1969. He was found dead in the bath of his Paris apartment

and died from heart problems, aggravated by alcohol. A coroner recorded a verdict of death by natural causes and his grave at the city's Pere Lachaise ceremony has become a shrine for fans.

Sumar uman

"We know he was at Florida State University for a period of time and he did some acting when he was there," Norman added.The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had been rejected by Florida State University.Morrison is seen quizzing a college administrator on why he was refused.Morrison was notorious for his wild lifestyle - and was accused of exposing himself and simulating a sex act at a Miami concert in 1969.Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.How come my parents and the state and the university didn't look ahead?"

Sumar automat bayes cu eliminare stopwords si lemantizare

Lost Doors frontman movie found

Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.

The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had

been rejected by Florida State University."But what happened?How come my parents and the state and the university didn't look ahead?"he is seen asking.

"It's incredible.He's so clean cut and soft-spoken," said Florida state archivist Jody Norman."We know he

was at Florida State University for a period of time and he did some acting when he was there," Norman added.He was found dead in the bath of his Paris apartment and died from heart problems, aggravated by alcohol.

TF-IDF

Interactiunea cea mai interesant mi-sa parut la marirea numarului de fraze selectat pentru sumar. Scorul rouge tinde sa crasca, fapt ce denota ca sunt surprinse mai multe cuvinte in sumar, insa scorul bleu scade ce inseamna ca sumarul nu este asa de concis, precum ne-am astepta. De exemplu:

Articol din entertainment

Lost Doors frontman movie found

Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.

The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had

been rejected by Florida State University. Morrison is seen quizzing a college administrator on why he was refused. "But what happened? How come my parents and the state and the university didn't look ahead?" he

is seen asking.

"It's incredible. He's so clean cut and soft-spoken," said Florida state archivist Jody Norman. "We know he was at Florida State University for a period of time and he did some acting when he was there," Norman added. The Doors were one of the most influential bands of the 1960s, with hits including Light My Fire and Riders On The Storm. Morrison was notorious for his wild lifestyle - and was accused of exposing himself

and simulating a sex act at a Miami concert in 1969. He was found dead in the bath of his Paris apartment

and died from heart problems, aggravated by alcohol. A coroner recorded a verdict of death by natural causes and his grave at the city's Pere Lachaise ceremony has become a shrine for fans.

Sumar uman

"We know he was at Florida State University for a period of time and he did some acting when he was there," Norman added.The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had been rejected by Florida State University.Morrison is seen quizzing a college administrator on why he was refused.Morrison was notorious for his wild lifestyle - and was accused of exposing himself and simulating a sex act at a Miami concert in 1969.Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.How come my parents and the state and the university didn't look ahead?"

Sumar automat cu sent_count = 5

Lost Doors frontman movie found

Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.

The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had been rejected by Florida State University.He's so clean cut and soft-spoken," said Florida state archivist Jody Norman.Morrison was notorious for his wild lifestyle - and was accused of exposing himself and simulating a sex act at a Miami concert in 1969.A coroner recorded a verdict of death by natural causes and his grave at the city's Pere Lachaise ceremony has become a shrine for fans.

Sumar automat cu sent_count = 3

Lost Doors frontman movie found

Historians in Florida have discovered a 40-year-old clip of a clean-cut Jim Morrison appearing in a promotional film for his university.

The 1964 film shows the Doors frontman, who died aged 27 in 1971, playing the part of a young man who had been rejected by Florida State University.A coroner recorded a verdict of death by natural causes and his grave at the city's Pere Lachaise ceremony has become a shrine for fans.