abelardo\_g\_fournier

## **README**

This is a script that estimates the similarity between documents that are taken two by two from a larger collection. The algorithm calculates the average frequencies of the words used within the collection of all the documents, and compares these values with the local frequencies of the words within each document individually. The deviations / fluctuations become then the parameters used to calculate a quantitative value of semantic proximity.

## More information:

http://scikit-learn.org/stable/modules/feature\_extraction.html#text-featureextraction (http://scikit-learn.org/stable/modules/feature\_extraction.html#textfeature-extraction) (Proximus NV » Level 3 Communications, Inc. » GitHub,

FEATURE EXTRACTION: MOST FREQUENT WORDS

An\_Ethnography\_of\_Error.txt

-- error(13) human(12) problem(11) driverless(9) car(9)

Computing\_War\_Narratives\_The\_Hamlet\_Evaluation\_System\_in\_Vietman.txt -- hamlet(22) hes(20) data(16) war(12) vietnam(12)

Elusive\_Borders\_Aesthetic\_Perpectualization\_The\_Space\_Time\_of\_Metadata.txt -- space(21) data(18) cables(15) metadata(14) tor(12)

From\_Page\_Rank\_to\_Rankbrain.txt

-- search(28) google(22) page(17) users(13) algorithms(11)

Machine\_Listening.txt

-- human(28) wavenet(22) ernst(18) machine(16) listening(16)

Machine\_Pedagogies.txt

-- images(16) human(15) freire(15) oppressor(13) learning(13)

Participation\_in\_Infrastructures.txt

-- systems(20) electricity(19) demand(13) time(12) resources(11)

Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt -- data(26) recognition(11) digital(11) capture(11) perception(9)

Relearn\_to\_Read\_Speed\_Readers.txt

-- reading(39) speed(35) reader(13) readers(12) new(12)

Resolution\_Theory.txt

-- resolution(66) term(17) word(15) resolutions(15) formal(14)

Testing\_Texting\_South\_a\_Political\_Fiction.txt

-- language(18) south(17) political(17) fictions(13) 2015(11)

The\_Cultural\_Politics\_of\_Information\_and\_of\_Debt.txt

-- information(20) debt(19) communication(19) terranova(15) message(14)

The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt -- language(14) communication(14) reality(11) thought(10) medium(10)

The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt

-- network(21) ip(15) tcp(13) internet(13) protocol(12)

Unmaking\_Screens\_a\_Genealogy\_of\_the\_Mineral\_Vision.txt -- vision(13) media(11) land(10) earth(10) practices(7)

CALCULATED SIMILARITY VALUE BETWEEN DOCUMENTS

An\_Ethnography\_of\_Error.txt

- 0.88 Mochine\_lespogles\_txt

- 0.88 Mochine\_lespogles\_txt

- 0.87 Mochine\_listering\_txt

- 0.867 From\_Poge\_Rank\_to\_Rankbroin.txt

- 0.867 From\_Poge\_Rank\_to\_Rankbroin.txt

- 0.868 Fasting\_Texting\_South\_APOlitical\_liction.txt

- 0.868 Fasting\_Texting\_South\_APOlitical\_liction.txt

- 0.869 Fasting\_Texting\_South\_APOlitical\_liction.txt

- 0.869 Fasting\_Texting\_South\_APOlitical\_liction.txt

- 0.869 Fasting\_Texting\_South\_APOlitical\_liction.txt

- 0.869 Taxing\_Texting\_txt

- 0.869 Taxing\_Texting\_South\_APOlitical\_liction\_Texting\_txt

- 0.874 Filest\_Nonders\_Aresthict\_Perpectual\_liction\_The\_Spoce\_Time\_of\_Metadota\_txt

- 0.868 Computing\_Non\_Norrorities\_The\_Hamlet\_Evaluation\_System\_in\_Vietnon.txt

- 0.864 Tax\_Outlurol\_Politics\_of\_Information\_and\_of\_Debt.txt

Computing\_Mar\_Norrotives\_The\_Hamlet\_Evaluation\_System\_in\_Vietnan.txt
- 0.083 Testing\_Texting\_South\_a\_Political\_Fiction.txt
- 0.078 Pattern-Recognition\_cross-Bodies-and-Rachines-by-Anarchival-Means.txt
- 0.079 Detach-Recognition\_cross-Bodies-and-Rachines-by-Anarchival-Means.txt
- 0.079 Demaking\_Screens\_Genealogy\_of\_the\_Mineral\_Vision.txt
- 0.079 Participrotion\_in\_Informstructures.txt

-- 0.824 Resolution\_Theory.txt

Elusive\_Borders\_Aesthetic\_Perpectualization\_The\_Space\_Time\_of\_Metadato.txt
-- 0.186 Participation\_in\_Infrastructures.txt
-- 0.185 The\_Stupid\_letwork\_thot\_we\_fono\_and\_Love.txt
-- 0.183 Pattern-Recognition\_coross-Bootles\_and-Machines-by-Anarchival-Means.txt
-- 0.827 The\_Cultural\_Politics\_of\_Information\_and\_of\_Debt.txt
-- 0.828 The\_Signification\_Communication\_Question\_Some\_Initial\_Remrks.txt
-- 0.837 Stracking\_Screens\_o\_deneology\_of\_the\_Mineral\_vision.txt
-- 0.808 The\_Diago\_Rank\_to\_Roundardin\_txt
-- 0.809 Too\_proje\_Rank\_to\_Roundardin\_txt
-- 0.809 Mechine\_Listering\_txt

-- 0.047 An\_Ethnography\_of\_Error.txt

Frow\_Dogw\_Rank\_to\_RankPrain.txt

-- 0.138 Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

-- 0.091 The\_Stupid\_Network\_thot\_me\_Know\_and\_Love.txt

-- 0.087 Nothine\_Pedagogies.txt

-- 0.087 Testine\_Pedagogies.txt

-- 0.087 Testing\_Testing\_South\_apolitical\_Fiction.txt

-- 0.085 Mochine\_Isterning\_txt

-- 0.085 Testing\_T

- e, vew Resolution, Incory.xxt

Mochine, Listering, txt
- 0.139 McChine, Pedgogles, txt
- 0.139 McChine, Pedgogles, txt
- 0.808 Festing, Texting, South, a, Political\_Fiction.txt
- 0.808 Festing\_Texting\_South, a, Political\_Fiction.txt
- 0.808 Angletongraphy, of Foron.txt
- 0.808 Angleton, Long, and Jenor.tx
- 0.808 Angleton, Long, and Jenor.tx
- 0.808 Foreign, Company, and Long, and Long, Long, and Long, Long, and Long, an

-- 0.836 Computing\_Mor\_Morratives\_The\_Momlet\_Evaluation\_System\_in\_Vietmon.txt
Monthine\_Pedagogies.txt
-- 0.139 Monthine\_Listening.txt
-- 0.088 An\_Ethnography.of\_Error.txt
-- 0.088 Forton\_Dege\_Ender\_to\_Montherin.txt
-- 0.088 Pottern-Recognition-across-Boolies-and-Mochines-by-Anarchival-Means.txt
-- 0.073 Inmosting\_Screens\_c\_denelogy\_of\_the\_Mineral\_Vision.txt
-- 0.074 Inmosting\_Screens\_c\_denelogy\_of\_the\_Mineral\_Vision.txt
-- 0.075 Relevent\_to\_tead\_Specific\_Moder\_txt
-- 0.075 Relevent\_to\_End\_Specific\_Moder\_txt
-- 0.075 Relevent\_txt
-

- 0.045 Machine\_Pedagogles.txt

Pattern-Recognition-ocross-Bodies-and-Machines-by-Anarchival-Means.txt

- 0.138 Fram\_Page\_Rank\_to\_Rankhorin.txt

- 0.138 Fram\_Page\_Rank\_to\_Rankhorin.txt

- 0.138 Fram\_Dage\_Rank\_to\_Rankhorin.txt

- 0.097 Relearn\_to\_Read\_Speed\_Readers.txt

- 0.097 Relearn\_to\_Read\_Speed\_Readers.txt

- 0.097 Tels\_inn\_Texting\_South\_o\_Political\_Fiction.txt

- 0.097 Tels\_inn\_Texting\_touth\_o\_Political\_Fiction.txt

- 0.097 Tels\_inn\_fication\_Commanication\_Question\_Some\_Initial\_Remarks.txt

- 0.098 Machine\_Listening.txt

- 0.089 Proliticyprion\_in\_Infrastructures.txt

- 0.089 Proliticyprion\_in\_Infrastructures.

- 0.097 Computin\_Bor\_Harratives\_The\_Manlet\_Evaluation\_System\_in\_Vietnon.txt

- 0.097 Computin\_Bor\_Harratives\_The\_Manlet\_Evaluation\_System\_in\_Vietnon.txt

- 0.096 The\_Studion\_Theory\_txt

- 0.096 The\_Studion\_Theory\_txt

- 0.096 The\_Studion\_Theory\_txt

- 0.096 The\_Studion\_Theory\_txt

- 0.096 The\_Studion\_For\_or\_txt

-- 0.061 An\_Ethnography\_of\_Error.txt

Relearn\_to\_Read\_Speed\_Readers.txt
-- 0.125 Testing\_Testing\_South\_political\_Fiction.txt
-- 0.079 Testing\_Testing\_South\_political\_Fiction.txt
-- 0.079 Testine\_Instaining\_txt
-- 0.083 Monther\_Listerining\_txt
-- 0.0827 The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt
-- 0.0827 The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt
-- 0.0839 The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt
-- 0.0830 The\_Signification\_Communication\_Question\_Signification\_txt
-- 0.072 Participation\_In\_Infortructures.txt
-- 0.072 Participation\_In\_Infortructures.txt
-- 0.072 Testinography\_of\_Error.txt
-- 0.0630 The\_Signification\_Communication\_Signification\_Signification\_txt
-- 0.0630 The\_Cimpagneyof\_Of\_Error.txt
-- 0.0630 The\_Cimpa

Resolution\_Theory.txt

- 0.071 The \_Signification\_Communication\_Question\_Some\_Initial\_Remorks.txt

- 0.072 The \_Signification\_Communication\_Question\_Some\_Initial\_Remorks.txt

- 0.060 Pattern Recognition-ocross-Bodies-and-Machines-by-Amarchival-Mems.txt

- 0.061 Descripts Recognition\_Initial\_Remorks\_Communication\_Communicat

The Cultural Politics, of Information, and, of Debt.txt

- 0.164 The Signification, Communication, Question, Some\_Initial\_Remarks.txt

- 0.164 The Signification, Communication, Question, Some\_Initial\_Remarks.txt

- 0.093 The Signification, Learning, at the Love.txt

- 0.093 The Signification, Learning, at the Love.txt

- 0.095 Machine\_Listering, It.

- 0.095 Machine\_Listering, It.

- 0.095 Machine\_Listering, It.

- 0.097 Machine\_Listering, Infrastructures, txt

The Signification Communication Question Some Initial Remarks txt

- 0.164 The Cultural Politics of Information and of Debt.txt

- 0.134 Testing Texting South, Political Fistion.txt

- 0.115 Machine Listening.txt

- 0.115 Machine Listening.txt

- 0.022 Pattern Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

- 0.035 The Signiful Methork.tbot.me.Know.and.Love.txt

- 0.035 The Signiful Methork.tbot.me.Know.and.Love.txt

The Stupid Network that we Now, and Love txt

- 0.127 Participation, in. Infrastructures.txt

- 0.126 Teurise Borders. Asstratic Perpectualization. The Space\_Time\_of\_Metadata.txt

- 0.195 Teurise\_Borders. Asstratic\_Perpectualization. The Space\_Time\_of\_Metadata.txt

- 0.195 Teurise\_Borders. Teurise\_Teurise\_Borders.txt

- 0.091 From\_Dogs\_Bonit. to, Bondheriani.txt

- 0.095 The Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt

- 0.082 Ummoking\_Screens\_d\_Geneelogy of.the Mineral\_Vision.txt

- 0.097 Testing\_Texting\_South\_a\_Political\_Fiction.txt

- 0.079 Nachine\_Listering\_txt

uchtne\_tistening.txt Vattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt Welearn\_to\_Read\_Speed\_Readers.txt Un Ethnoarophy of Error.txt

c\_Perpectualization\_The\_Space\_Time\_of\_Metadata.txt

## RECOMMENDED READINGS FLOW

If you enjoyed reading An\_Ethnography\_of\_Error.txt, you might like also:

-- Machine\_Pedagogies.txt

-- Machine\_Listening.txt

If you enjoyed reading Computing\_War\_Narratives\_The\_Hamlet\_Evaluation\_System\_in\_Vietman.txt, you might like also:

abelardo g fournier

-- Testing\_Texting\_South\_a\_Political\_Fiction.txt

-- Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

If you enjoyed reading Elusive\_Borders\_Aesthetic\_Perpectualization\_The\_Space\_Time\_of\_Metadata.txt, you might like also:

-- Participation\_in\_Infrastructures.txt

-- The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt

If you enjoyed reading From\_Page\_Rank\_to\_Rankbrain.txt, you might like also:

-- Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

-- The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt

If you enjoyed reading Machine\_Listening.txt, you might like also:

-- Machine\_Pedagogies.txt

 $\hbox{\it -- The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt}$ 

If you enjoyed reading Machine\_Pedagogies.txt, you might like also:

-- Machine\_Listening.txt

-- An\_Ethnography\_of\_Error.txt

If you enjoyed reading Participation\_in\_Infrastructures.txt, you might like also:

-- Unmaking\_Screens\_a\_Genealogy\_of\_the\_Mineral\_Vision.txt

-- The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt

-- Testing\_Texting\_South\_a\_Political\_Fiction.txt

If you enjoyed reading Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt, you might like also:

-- From\_Page\_Rank\_to\_Rankbrain.txt

-- Elusive\_Borders\_Aesthetic\_Perpectualization\_The\_Space\_Time\_of\_Metadata.txt

If you enjoyed reading Relearn\_to\_Read\_Speed\_Readers.txt, you might like also:

-- Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

If you enjoyed reading Resolution\_Theory.txt, you might like also: -- The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt

-- Pattern-Recognition-across-Bodies-and-Machines-by-Anarchival-Means.txt

If you enjoyed reading Testing\_Texting\_South\_a\_Political\_Fiction.txt, you might like also:

 $-- The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt$ -- Relearn\_to\_Read\_Speed\_Readers.txt

If you enjoyed reading The\_Cultural\_Politics\_of\_Information\_and\_of\_Debt.txt, you might like also: -- The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt

-- The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt

If you enjoyed reading The\_Signification\_Communication\_Question\_Some\_Initial\_Remarks.txt, you might like also:

-- The\_Cultural\_Politics\_of\_Information\_and\_of\_Debt.txt

-- Testing\_Texting\_South\_a\_Political\_Fiction.txt

If you enjoyed reading The\_Stupid\_Network\_that\_we\_Know\_and\_Love.txt, you might like also:

-- Participation\_in\_Infrastructures.txt

-- Elusive\_Borders\_Aesthetic\_Perpectualization\_The\_Space\_Time\_of\_Metadata.txt

If you enjoyed reading Unmaking\_Screens\_a\_Genealogy\_of\_the\_Mineral\_Vision.txt, you might like also: -- Participation\_in\_Infrastructures.txt

-- Testing\_Texting\_South\_a\_Political\_Fiction.txt