

Classifying Documents According to Locational Relevance

Ivo Anastácio, Bruno Martins, Pável Calado

EPIA — October 15, 2009



Context

1

cinema em

MAGAZINE EM CARTAZ ESTREMA CINEGRAFIA TRAILERS GALERIAS PASSATIMOS

EM DESTAQUE BOSSEROS * e-CINEMA * FILMES * PERSONAS * SFILMES e-CINEMA

e-CINEMA [divulgar](#) [inspirar](#) [compartilhar](#)

e-Cinema: Corrupção, tramas políticas e cinema francês

A semana da 7ª Arte traz-nos temas de corrupção, uma vida de guerra eletrónica, um filme baseado na obra de Bret Easton Ellis e muito cinema francês.

Esta semana no «e-Cinema» filmes de «**Orgulho e Prejuízo**», um filme sobre temas de corrupção na política com Edward Norton e Colin Farrell, «**Blow**», um «biopic» sobre a figura incontrolável da política italiana, e «**Os Informadores**», uma película baseada na obra de Bret Easton Ellis.

Fora de circuito, mergulhamos no 10º Festival do Cinema Francês.

Nos recomendamos, salientamos até uma menção recente para lembrarmos os sucessos de **Alain Resnais** em «**Crash**».

Vêja a edição desta semana do «e-Cinema» [link](#) [Clara Mota](#) - 08-20-2009

RELACIONADO [Comentários](#)

FILMES PERSONAS

2

BLITZ Filipa de BLITZ e sua HomePage [download](#) [optima](#) [optima](#) [tema](#)

Home Notícias Fórum Agenda Festivais Emprego Artistas de A a Z Classificados Games

BLITZ 0859 Comos Esperto Esperto

Green Day ao vivo no Pavilhão Atlântico [texto + fotos + alinhamento]

Furacão Green Day arrasou com Lisboa esta noite. Muitos fãs em palco ajudaram a fazer a festa que, apesar de tudo, não durou os três horas prometidas. Consulte aqui o alinhamento.

A parada era alta, as expectativas estavam lá bem em cima, infacionadas por anos de espera e pelo facto de a banda neste período se ter tornado um dos maiores fenómenos rock do planeta. E que faziam os Green Day now anos depois de um concerto no Coliseu dos Recreios? Amassam com Lisboa. A nota só não terá sido máxima porque os músicos ficaram a um Pavilhão Atlântico perto de esgotar gostariam, com certeza, de ter ouvido "When I Come Around" ou "Wake Me Up When September Ends" e de terem sido brindados com as três horas de actuación prometidas (foz-se pelas duas horas e quinze minutos).

3

lifecooler O Guia da Boa Vida O site de turismo mais visitado

[Pesquise no Lifecooler como no Google](#)

[Quer receber a e-Letter?](#)

Restaurante Alma - Lisboa

Restaurante Alma - Lisboa
Cozinha de autor com o cunho do chef Henrique Sa Pessoa
Ana Abrachens

Com menos de um ano de existência, o Alma é um exemplo de sucesso, com a casa sempre cheia. O ambiente, a decoração, o serviço e a cozinha apontam em uníssono para a chave do êxito: o encontro harmonioso entre a mais alta qualidade e um acolhimento primoroso sem descurar o conforto.

Depois de sete anos de trabalho em conjunto, Henrique Sa Pessoa e o seu sub-chef, Daniel Costa, estreiam-se na aventura de abrir um restaurante próprio. Uma aventura sem muitos perigos

Context
Solution
Evaluation
Results
Conclusions

Objective

Estimate how important is the geographic content of a given document, i.e., the **locational relevance** of the document.

- Several applications:
 - Geographic Information Retrieval
 - Contextual Advertising
 - many more ...

Context

The screenshot displays the eDreams website interface. At the top, there is a navigation bar with 'Destinos', 'Eventos', 'Noite', and 'Restaurantes'. Below this is a banner for 'VOOS a partir de 20€'. The main content area features a section titled 'Restaurantes em Florença' with a sub-header 'Restaurants em Florença'. The text describes the gastronomy of Tuscany and lists several restaurants: Acqua al 2, Bocca Negra, and Borgo San Jacopo. To the right, there is a flight search form with fields for 'De:', 'A:', 'Partida:', and 'Regresso:', and a 'Procurar' button. Below the search form, a search results box is highlighted with a red border, showing results for '11ª Feira Gastronomia' and 'Restaurante Clube da Gula'.

Restaurants em Florença

A gastronomia da **Toscana**, onde Florença está situada, é saborosa e original, mas é simples, baseada na qualidade dos produtos da região e em um tratamento sofisticado dos alimentos. Uma boa comida regional começa com uma entrada de embutidos e *crostini*, que são fatias de pão tostadas com patês. O primeiro prato pode ser uma sopa de verduras da estação, condimentada com queijo parmesão ralado e azeite de oliva. Entre as pastas, sempre frescas e preparadas de forma artesanal, destaque para a *pappardelle* com molho de lebre. Não deixe de provar os deliciosos *pici*, uma espécie de *spaghetti* grosso que se serve com vários tipos de molhos.

Aproveite o melhor da cozinha italiana e toscana nos restaurantes de Florença. Conheça algumas opções em nosso guia.

[Guia de Florença](#) | [A noite de Florença](#)

Acqua al 2

Prove o cardápio com novas leituras da gastronomia italiana no Acqua al 2. Cozinha criativa e inovadora

Bocca Negra

O restaurante Bocca Negra é o lugar ideal para um jantar romântico em Florença, com ambiente acolhedor e pratos criativos

Borgo San Jacopo

Decoração dos anos 50 e ambiente elegante. Reserve uma mesa ao lado do rio Arno e desfrute de uma agradável noite no Borgo San Jacopo

11ª Feira Gastronomia
De 21 a 30 de Agosto de 2009 Cozinha Portuguesa em Vila do Conde
gastronomia.vconde.org

Jantares de Luxo a Dois
Surpreenda quem gosta com um menu requintado. Ambiente elegante.
www.pai.pt/Restaurantebairro

Restaurante Clube da Gula
Sardinhada S. João dia 23 Junho Antiga trav. do Rosário ao HGSA
www.clubeda-gula.pt

Context

Solution

Evaluation

Results

Conclusions

Solution

Context

Solution

Evaluation

Results

Conclusions

- Classify documents as **global** or **local**
- Use the probability estimates as the locational relevance
- Narrow geographic contexts are more interesting than broad geographic contexts
- Two types of features:
 - Thematic
 - Geographic

Thematic Features

Context

Solution

Evaluation

Results

Conclusions

Capture the thematic aspects of the local and global classes that are encoded in the text, as well as on the URL.

- **Textual features**

- lowercasing, stemming, and stop-word removal
- TF.IDF of all words
- TF.IDF of key words (Yahoo! Term Extraction)

- **URL features**

- TF.IDF of URL n -grams, with n between 4 and 8

Geographic Features

Characterize the geographic context of the document.

- **Simple Locative features**

- Counts for place references in text
- More than twenty features
 - Total locations
 - Unique Locations
 - Total cities
 - ...

- **High-Level Locative features**

- Considers the results from two geographic scope assignment algorithms (Web-a-Where, GIPSY)
- Area of the geographic scope
- Confidence in geographic scope assignment

Technicalities

Context

Solution

Evaluation

Results

Conclusions

- Support Vector Machines (Weka + LIBSVM)
 - *Lin, H., Lin, C., and Weng, R. (2007). A note on platt's probabilistic outputs for support vector machines. Machine Learning, 68(3), pp. 267 - 276*
- Yahoo! Term Extraction - Extracts keywords
- Yahoo! Placemaker - Recognizes place references
- Yahoo! GeoPlanet - Provides geographical information

Evaluation

Context

Solution

Evaluation

Results

Conclusions

- Two datasets with local and global classes equally split
 - 8.000 pages from the Open Directory Project
 - 100 news pages from local news aggregator Topix.com
- Precision, Recall, F_1 , and Accuracy
- 10-fold cross validation

Results

ODP dataset

Experiment	Precision		Recall		F-Measure		Error	Accuracy
	Local	Global	Local	Global	Local	Global		
TW	0.81	0.79	0.79	0.81	0.8	0.8	19.86	80.14
TK	0.74	0.74	0.74	0.74	0.74	0.74	25.86	74.14
TWK	0.82	0.8	0.8	0.82	0.81	0.81	19.09	80.92
U	0.66	0.69	0.72	0.63	0.69	0.66	32.47	67.53
LS	0.78	0.86	0.88	0.74	0.82	0.8	18.85	81.15
LH	0.56	0.69	0.87	0.3	0.68	0.42	41.31	58.69
LSH	0.58	0.82	0.93	0.31	0.71	0.45	37.58	62.42
T+L	0.89	0.91	0.91	0.88	0.9	0.89	10.43	89.57
T+L+U	0.89	0.91	0.91	0.89	0.9	0.9	10.15	89.85

- Simple locative features are the most discriminative
- Combining textual and locative features results in improvements

Results

Topix dataset

Experiment	Precision		Recall		F-Measure		Error	Accuracy
	Local	Global	Local	Global	Local	Global		
TW	0.53	1	1	0.1	0.69	0.18	45	55
TK	0.71	0.61	0.48	0.8	0.57	0.69	36	64
TWK	0.54	1	1	0.16	0.7	0.28	42	58
U	0.45	0.39	0.6	0.26	0.51	0.31	57	43
LS	0.63	0.6	0.54	0.68	0.58	0.64	39	61
LH	0.54	0.51	0.24	0.8	0.33	0.62	48	52
LSH	0.67	0.62	0.56	0.72	0.61	0.67	36	64
T+L	0.71	0.64	0.58	0.76	0.64	0.7	33	67
T+L+U	0.54	0.65	0.86	0.26	0.66	0.37	44	56

- Topix dataset presented more difficult challenges
- Combination of textual and locative features also results in improvements

Conclusions and Future Work

Context

Solution

Evaluation

Results

Conclusions

- Approach for locational relevance classification
- Empirical evaluation over two datasets
- A combination of thematic and geographic features obtained the best results

- Future Work
 - Explore information from incoming links
 - Evaluate impact of place reference resolution performance

The End

Context

Solution

Evaluation

Results

Conclusions

Thank You!

Questions?