

ASPECT CLUSTERING METHODS FOR SENTIMENT ANALYSIS

Francielle A. Vargas (MSc student)
Thiago A. S. Pardo

Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences
University of São Paulo



SENTIMENT ANALYSIS

- Pang et al. (2002)
 - Area in charge of identifying, extracting and summarizing subjective information in texts
- Opinion Mining (OM)
 - Natural Language Processing
 - Text Mining
 - *User Generated Content*

APPLICATIONS



e-bit

Em 03/03/2017

 Verificado

 10/10

 **Recomendo este produto**

Ótimo smartphone!!

Não tem do que reclamar sobre o iPhone 7. Simplesmente incrível, do design à todas as funcionalidades. Na verdade, tem uma coisa para reclamar sim, o preço absurdo. Infelizmente somos reféns desse tipo de prática aqui no BR.

Achou essa opinião útil?  58  15

APPLICATIONS



e-bit

Em 03/03/2017

Verificado

★★★★★ 10/10

Recomendo este produto

Ótim

Não t
verda
no BF

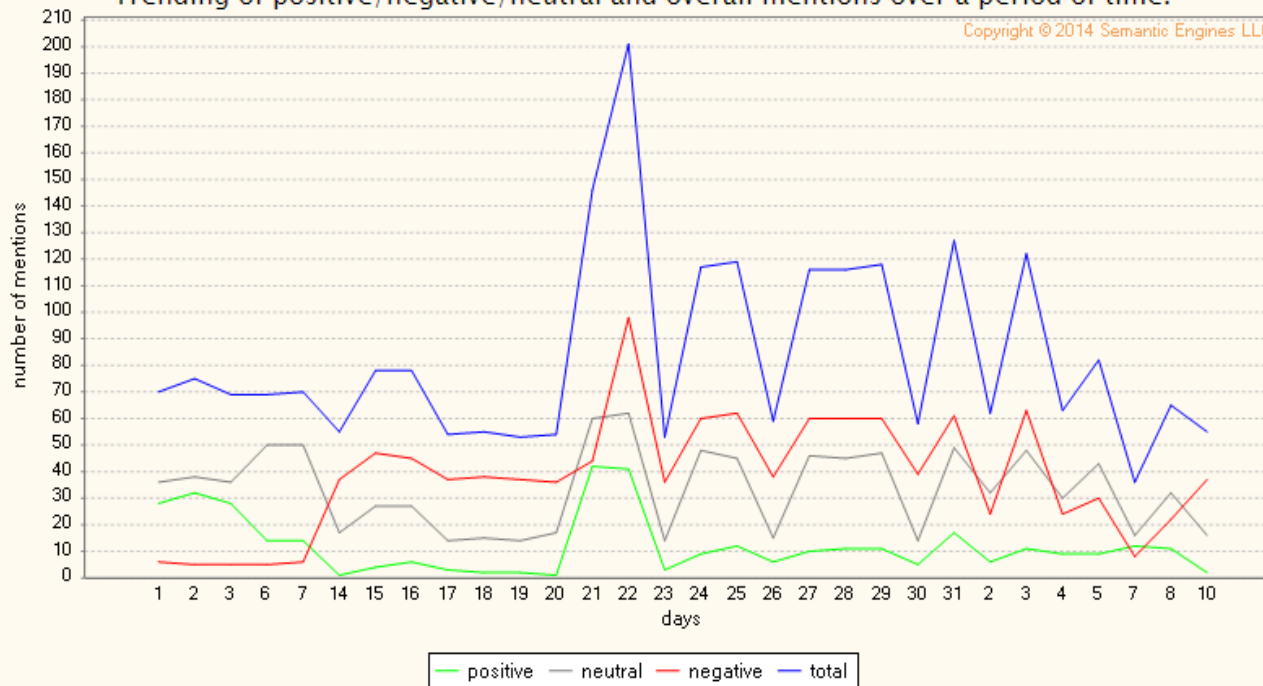
Acho

nalidades. Na
ipo de prática aqui

Sentiment Trend for: iPad

Trending of positive/negative/neutral and overall mentions over a period of time.

Copyright © 2014 Semantic Engines LLC



APPLICATIONS



e-bit

Em 03/03/2017

Verificado

★★★★★ 10/10

Recomendo este produto

Ótim

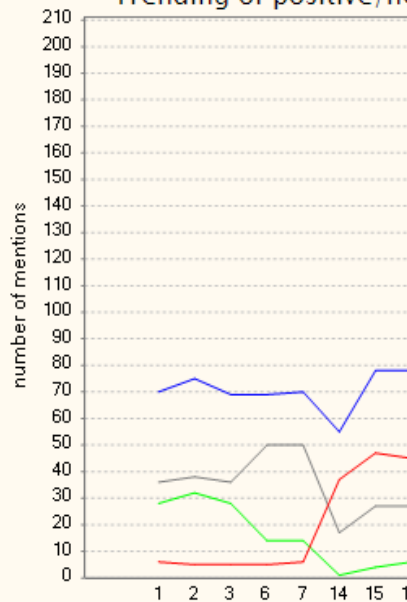
Não t
verda
no BF

Acho

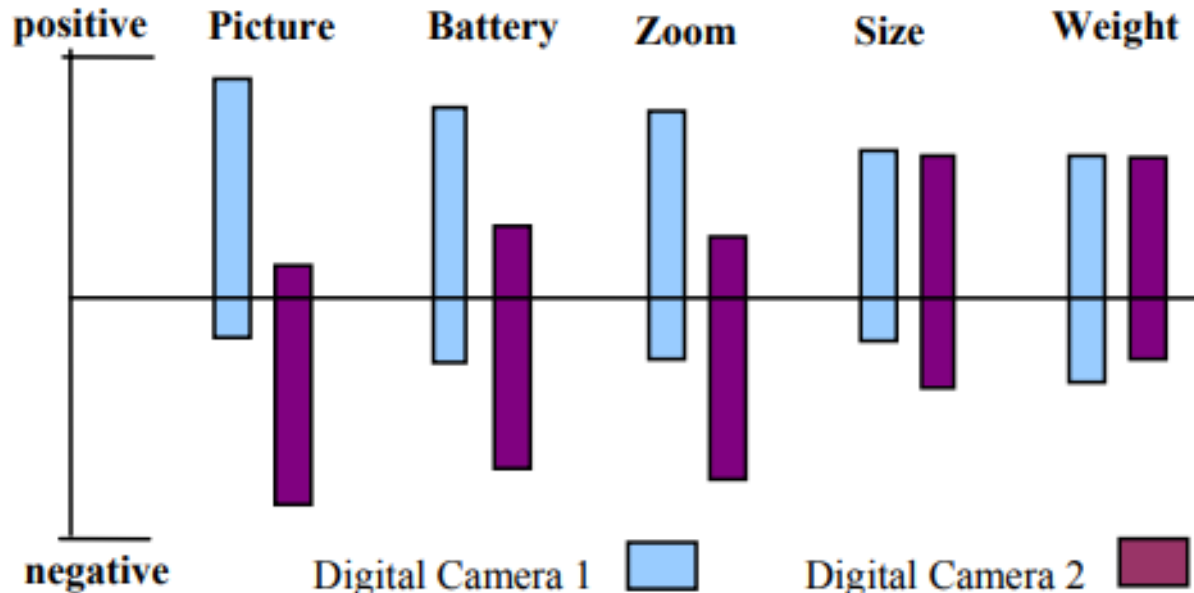
Sentiment Trend for: iPad

Trending of positive/negative/neutral and overall mentions over a period of time.

Copyright © 2014 Semantic Engines LLC



nalidades. Na
ipo de prática aqui



APPLICATIONS

Duração da Bateria

Positivo: 7 avaliações

- é muito boa
- está com melhor duração

...

Negativo: 9 avaliações

- tem baixa duração com gps ligado
- acaba muito rápido.

...

Câmera

Positivo: 9 avaliações

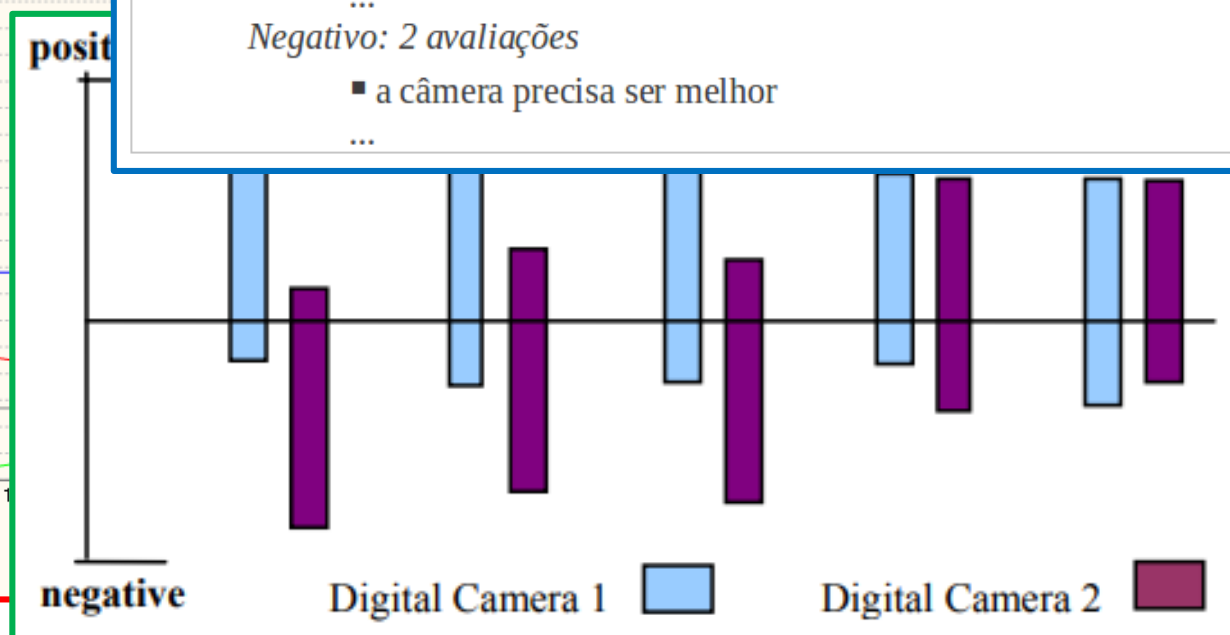
- a câmera frontal e traseira são excelentes
- suas fotos tem boa qualidade
- é ótima

...

Negativo: 2 avaliações

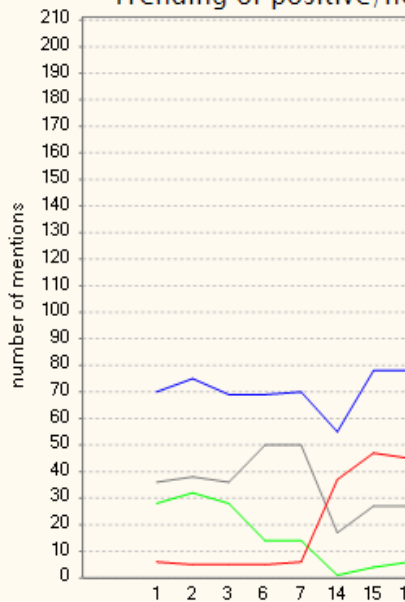
- a câmera precisa ser melhor

...



Sentiment

Trending of positive/negative/



e-bit
Em 03/03/2017 Verificado

Ótim
Não t
verda
no BF
Acho

RELATED TASKS

- Opinion retrieval
- Opinion qualification
- Subjectivity classification
- Aspect identification and extraction
- Polarity classification
- Summarization
- Visualization
- Etc.

Varied challenges!

RELATED TASKS

- Opinion retrieval
- Opinion qualification
- Subjectivity classification
- **Aspect identification and extraction**
- Polarity classification
- Summarization
- Visualization
- Etc.



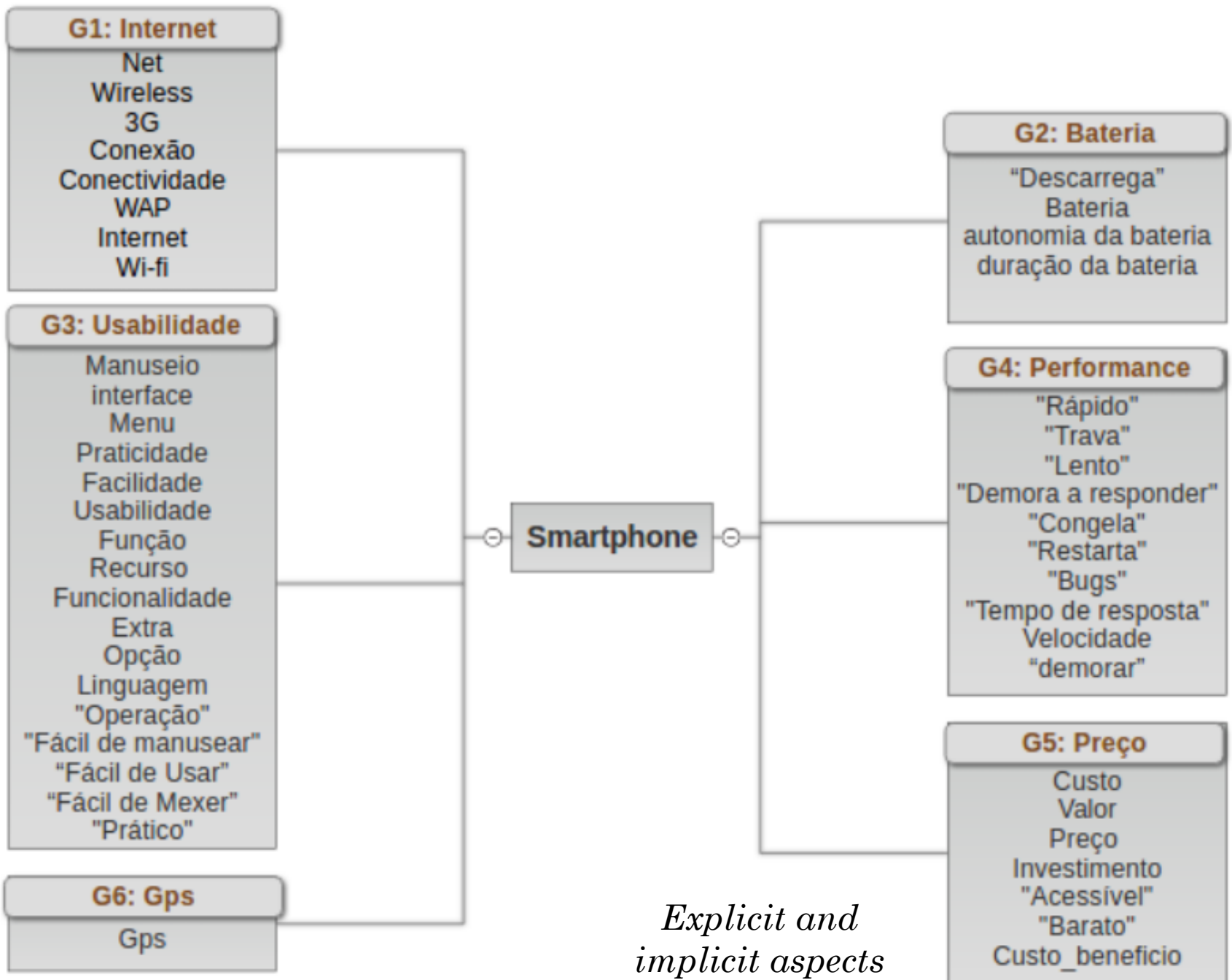
Task: aspect
clustering

Varied challenges!

EXAMPLES

O *iPhone 8 Plus* dispensa comentários: **aparelho** robusto, excelente **sistema operacional**, qualidade e **acabamento**...

Aparelho com fluidez nas aplicações, apanhei do **ios** no início e até hoje aprendo novas funcionalidades, mas realmente supera as expectativas, recomendo a todos que ainda não possuem, **design** simplesmente fantástico...



OUR GOAL

- To automatically identify the **different ways that users refer to the aspects** of interest
- To subsidize and improve *OM tools* for Portuguese
 - Aspect extraction (Balage Filho et al., 2014; Catharin and Feltrim, 2018)
 - Polarity classification (Avanço et al., 2016)
 - Opinion summarization (López Condori and Pardo, 2017)
 - Etc.

PREVIOUS APPROACHES

○ Machine learning

- Zhai et al. (2011): semi-supervised, English, products
- Zhang et al. (2011): semi-supervised, Chinese, products
- Abujbara et al. (2013): semi-supervised, Arabic, politics
- Zhou et al. (2015): supervised, English, restaurants
- Chen et al. (2016): semi-supervised, Chinese, products

○ Linguistic approaches: lexicons, thesauri and ontologies

- Patra et al. (2014): wordnet, English, products and restaurants
- Garcia et al. (2014): wordnet and wikipedia, English, products and restaurants

METHODS

- 4 linguistic methods inspired in the literature, in an incremental strategy
 - Synonym
 - Synonym + hypernym
 - Synonym + hypernym + meronym
 - Synonym + hypernym + meronym + coreference
- 1 statistical method
 - Word embeddings
- 1 enriched linguistic-based method (OpCluster-PT)
 - Synonym + hypernym + meronym + coreference
 - Causative, deverbal, diminutive/augmentative, foreignism and substring relations
 - Based on an empirical research (Vargas and Pardo, 2018)

TOOLS AND RESOURCES

- NILC embeddings (Hartmann et al., 2017)
- Lexical relations from Onto.PT (Oliveira, 2014)
- Coreference resolution by CORP (Fonseca et al., 2017)
- Other resources
 - Portuguese foreignism and deverbal iLteC lexicons (Janssen and Ferreira, 2007; Ferreira and Janssen, 2017)
 - List of diminutive/augmentative words

CORPUS

- Manually labeled data
 - ReLi (Freitas et al., 2012)
 - Buscapé (Hartmann et al., 2014)

Domain	Reviews	Tokens	Types
Book	60	35,771	1,577
Smartphone	60	6,077	1,496
Camera	60	3,887	1,060

CORPUS

- Manually labeled data
 - ReLi (Freitas et al., 2012)
 - Buscapé (Hartmann et al., 2014)

Domain	Reviews	Tokens	Types
Book	60	35,771	1,577
Smartphone	60	6,077	1,496
Camera	60	3,887	1,060

Domain	Book	Camera	Smartphone
total number of aspects	103	132	180
number of explicit aspects	91	109	142
number of implicit aspects	12	23	38
number of clusters	21	36	48

RESULTS

○ Precision

Methods	Book	Camera	Smartphone
synonymy	0.974	0.987	0.973
synonymy + is-a	0.916	0.967	0.940
synonymy + is-a + part-of	0.916	0.967	0.943
synonymy + is-a + part-of + coreference	0.945	0.963	0.953
word embeddings	0.953	0.962	0.956
opcluster-pt	0.925	0.933	0.947

- High number of unitary clusters

RESULTS

- Recall

Methods	Book	Camera	Smartphone
synonymy	0.231	0.281	0.296
synonymy + is-a	0.242	0.287	0.314
synonymy + is-a + part-of	0.242	0.287	0.310
synonymy + is-a + part-of + coreference	0.321	0.307	0.364
word embeddings	0.231	0.292	0.300
opcluster-pt	0.748	0.687	0.550

RESULTS

○ F-Measure

Methods	Book	Camera	Smartphone
synonymy	0.374	0.438	0.454
synonymy + is-a	0.383	0.442	0.471
synonymy + is-a + part-of	0.383	0.442	0.466
synonymy + is-a + part-of + coreference	0.480	0.466	0.527
word embeddings	0.372	0.448	0.457
opcluster-pt	0.827	0.792	0.702

- Better results for books: more prototypical / less domain specific aspects
 - Technical aspects are more difficult to find in the repositories

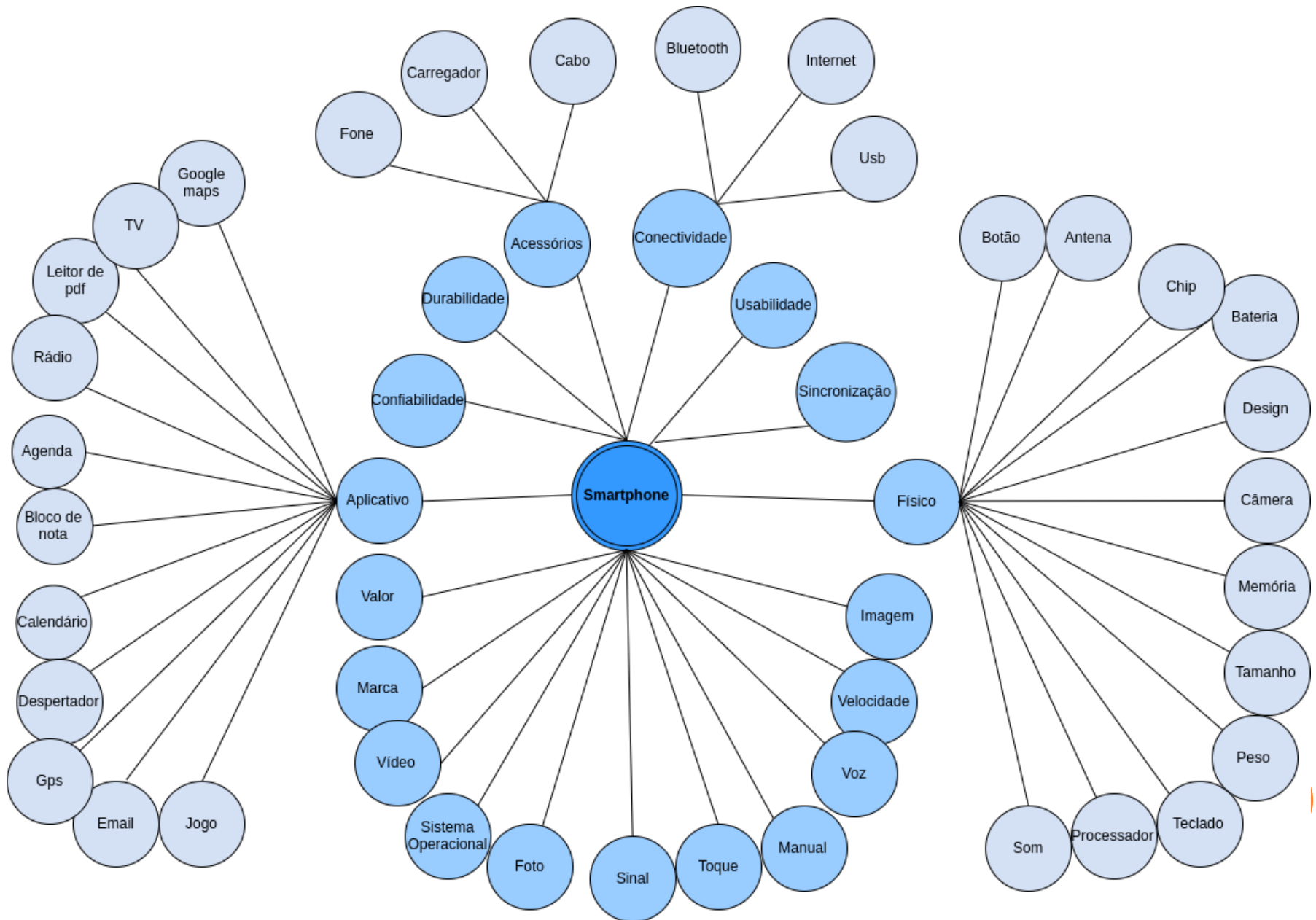
EXAMPLES: RELATIONS

- Hypernym: *câmera* ↔ *máquina, equipamento, produto*
- Synonym: *preço, custo* ↔ *valor*
- Meronym: *teclado* ↔ *tecla*
- Causative: *terminar* ↔ *fim*
- Diminutive: *livro* ↔ *livrinho*
- Deverbal construction: *manusear* ↔ *manuseio*
- Foreignism: *display* ↔ *tela*
- Substring: *imagem* ↔ *qualidade da imagem*
- Coreference: *garoto, edward, larry, bella, moleque* ↔ *personagem*

EXAMPLES: CLUSTERS

- A good cluster
 - {cost\benefit, price, value, investment, cheap}
- A mixed cluster
 - {enterprise, lg, nokia, sony, sony ericson, program, sign, system, model}

SIDE EFFECT: ASPECT ONTOLOGIES



FUTURE WORK

- To improve the results
 - Proper names
 - Wikipedia data
 - Specialized lexicons
 - Training of better embeddings
 - More informed methods: aspect modifiers

ASPECT CLUSTERING METHODS FOR SENTIMENT ANALYSIS

Thank you!

Our method, datasets and other related information are available at **OPINANDO project** website

<https://sites.google.com/icmc.usp.br/opinando/>

