

MediaViz: Agregação e visualização de conteúdos noticiosos online

Tiago Devezas¹ e Sérgio Nunes²

¹SAPO Labs/U.Porto

²INESC TEC e FEUP, U.Porto

Congresso de Humanidades Digitais em Portugal
Outubro 2015

Desafio

- Volume crescente de conteúdo online
 - Sites noticiosos
 - Blogues
 - Redes sociais
- Espalhado por múltiplas fontes
- Analisar e estudar este conteúdo através de um processo manual pode ser complexo e moroso

Oportunidade

- Capitalizar o uso de ferramentas informáticas para agilizar e automatizar a recolha e a análise
 - Trabalhar sobre um grande volume de dados
 - Visualizar rapidamente padrões e tendências

Oportunidade

- Aprofundar o conhecimento sobre os meios online
 - Melhorar o acesso à informação
 - Explorar diferentes tipos de interfaces
 - Investigar novos métodos de pesquisa

Abordagem

- Agregar e consolidar o conteúdo publicado por várias fontes num repositório centralizado
- Disponibilizar o acesso aos dados
- Explorar e interagir com os dados recolhidos através de ferramentas de visualização

Visualização

- As ferramentas visuais são excelentes para explorar grandes coleções de dados
- Tiram partido da percepção visual para ajudar a identificar tendências, padrões e ocorrências pouco usuais
- Auxiliam o processo de análise e compreensão

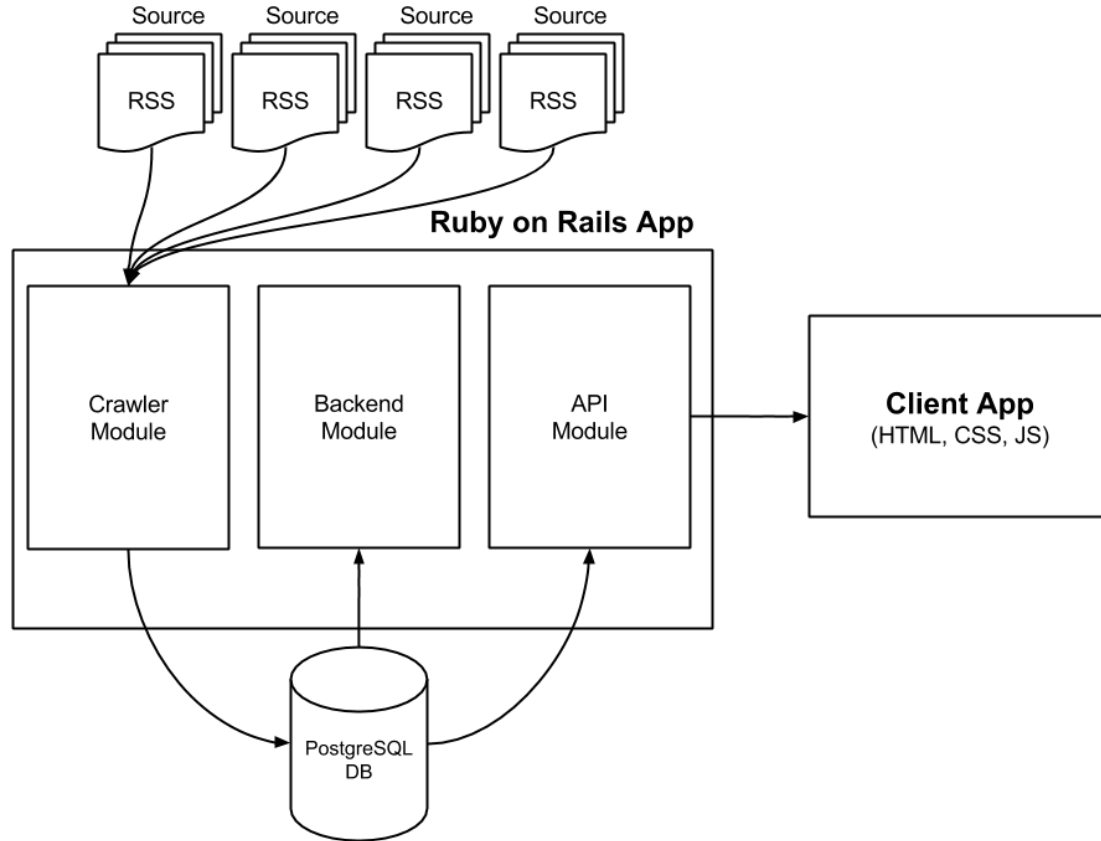
Trabalhos relacionados

- MediaCloud, BBC News Labs Juicer
 - Ambos fornecem APIs com dados recolhidos de milhares de fontes
 - Identificação de entidades, anotações semânticas, extração de texto, ...
 - Ferramentas de visualização

Porquê?

- Fontes, funcionalidades e dados que queríamos explorar em falta
 - Fontes portuguesas e blogues
 - Diferentes granularidades temporais
 - Partilhas no Twitter e Facebook por artigo
 - Cobertura geográfica ao nível dos distritos portugueses

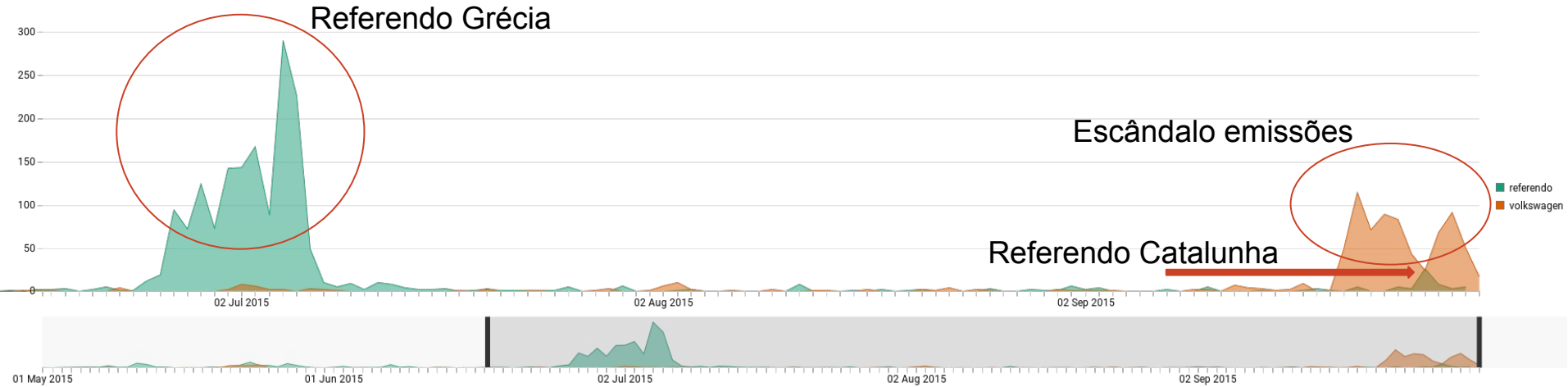
Plataforma MediaViz



Ferramentas de visualização

- Crónica

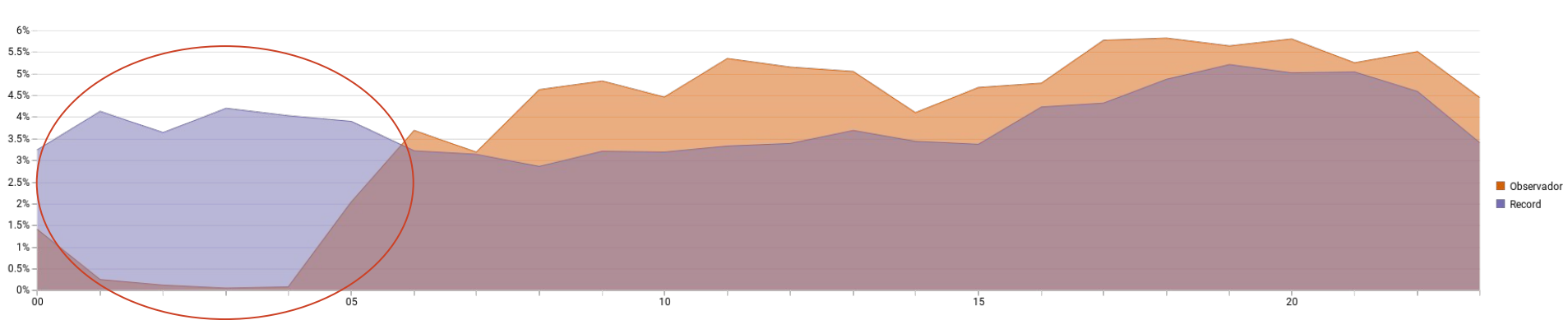
- Utilizadores podem introduzir vários termos de pesquisa e ver a cobertura respectiva ao longo do tempo
- Podem clicar nos pontos do gráfico para verem os artigos que motivaram a cobertura
- Permite ver os temas que estiveram ou estão sob o foco dos media, e de que forma essas dinâmicas mudam (com o tempo, de acordo com a fonte)



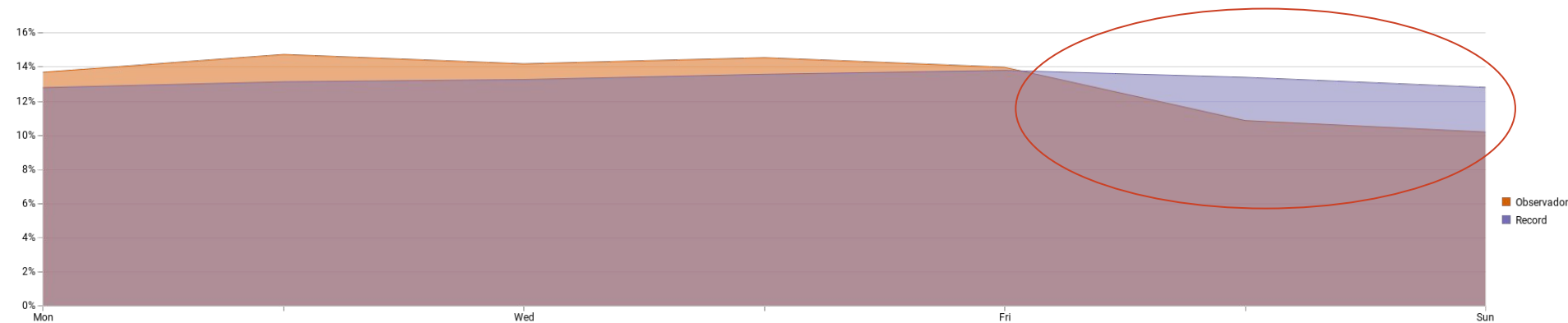
Volume de artigos publicados pelas fontes portuguesas com os termos “referendo” e “volkswagen”. Utilizadores podem clicar nos pontos do gráfico para verem os artigos associados e perceberem a razão para os picos na cobertura: o referendo à austeridade na Grécia e o escândalo das emissões da Volkswagen. O referendo à independência na Catalunha também é visível, embora os media lhe tenham conferido menor importância.

Fluxo

- Permite visualizar e comparar os padrões de publicação das fontes
- Diferentes granularidades temporais (hora, dia, semana, mês), intervalos definidos pelo utilizador
- Volume de artigos publicados, partilhas nas redes sociais



Percentagem de artigos publicados num ciclo de 24 horas pelas fontes “Observador” e “Record”. “Record” continua a publicar durante a madrugada, ao contrário do “Observador”.



Percentagem de artigos publicados num ciclo semanal pelas fontes “Observador” e “Record”. Novamente padrões distintos: no “Observador” há uma queda clara de publicações durante o fim de semana, enquanto que no “Record” a quebra é muito menos acentuada.

NewsMap

- Mapas do mundo e de Portugal
- Exibe o volume de artigos que mencionam cada país ou distrito português
- Co-ocorrências dos termos de pesquisa e o nome de cada região
- Cobertura geográfica das fontes online, ver quais as regiões mais associadas a diferentes termos (p.e. terrorismo)

Conclusão

- Grande quantidade de informação publicada pelos meios online
- Importante compreender o comportamento desta informação para desenvolver novas técnicas de acesso, interação e pesquisa
- Auxiliar investigadores, académicos e profissionais dos media através do uso de ferramentas informáticas

Conclusão

- Questões

- Que regiões (ao nível local e global) são alvo do foco dos media? Ou ignoradas?
- Que importância dão diferentes fontes a um mesmo tema?
- Esse grau de importância é reflectido nas redes sociais?
- Diferentes fontes apresentam padrões de publicação distintos?

Demo

<http://irlab.fe.up.pt/p/ht2015demo/>