

Social Web Engineering HackerSchool

Manuel Barreto Lima Reis, André Maurício Baltazar, João Alexandre

Relatório de Actividades

Resumo—Este relatório descreve o conjunto de actividades realizadas no contexto da tarefa de *Social Web Engineering promovida pela HackerSchool no âmbito da disciplina de Portefólio Pessoal III*. É apresentada uma descrição pormenorizada de cada uma das tarefas e actividades desempenhadas pelo grupo. Esta actividade tinha como principal objectivo a análise das publicações da *HackerSchool* nas redes sociais. Avaliámos estas publicações em relação ao seu sucesso e alcance e posteriormente gerámos um conjunto de recomendações que visam aumentar o sucesso de futuras publicações.

Palavras Chave—Actividades, Relatório, HackerSchool, Web, Redes Sociais, Estatísticas, Social Web Engineering.

porque não se fez esta referência no Introduction?

1 INTRODUÇÃO

ESTA actividade foi promovida pela *HackerSchool* e surgiu da necessidade que esta entidade tinha de melhorar a sua comunicação nas redes sociais. A actividade de *Social Web Engineering* consiste em recolher o conjunto de publicações de uma dada entidade numa rede social (neste caso o Facebook), analisá-las por forma a encontrar padrões entre características dessas publicações e o sucesso da mesmas (e.g, alcance, número de likes, etc) e posteriormente criar um conjunto de recomendações para que a entidade possa efectivamente melhorar a sua comunicação e atingir de forma mais eficaz o seu público-alvo.

Este relatório divide-se em duas partes, primeiro apresenta-se, na secção 2, a descrição das actividades inerentes à própria organização (i.e., comunicação, reuniões), e apresenta-se finalmente, na secção 3, as actividades directamente relacionadas com *Social Web Engineering*.

- Manuel Barreto Reis, nr. 69518,
E-mail: manuel.barreto.reis@tecnico.ulisboa.pt,
- André Maurício Baltazar, nr. 73137,
E-mail: andre.mauricio.baltazar@tecnico.ulisboa.pt,
- João Alexandre, nr. 73754,,
E-mail: joaopmalexandre@hotmail.com,
Instituto Superior Técnico, Universidade de Lisboa.

13 de Janeiro de 2015.

2 ACTIVIDADES ORGANIZACIONAIS

Esta secção visa descrever as actividades inerentes à realização de um trabalho conjunto, i.e., reuniões, contactos com entidades externas, etc.

2.1 Contactos com Entidade Promotora

Inicialmente cada um dos elementos do grupo teve uma reunião presencial com a representante da entidade promotora, a Inês Davim. Nesta reunião foi dada a cada um de nós qual a visão pretendida aquando da promoção da dita actividade.

Posteriormente, a Inês Davim, forneceu a cada um de nós os contactos dos restantes elementos do grupo para que pudéssemos dar início à realização da actividade.

Ao longo da actividade, foram feitos apenas alguns contactos pontuais por parte do grupo, via correio electrónico, com o intuito de pedir esclarecimentos sobre questões específicas inerentes à actividade.

Finalmente, foi enviado à Inês Davim um relatório conjunto com os resultados do trabalho efectuado.

2.2 Organização Interna

Toda a comunicação e coordenação entre o grupo foi realizada de forma não presen-

	ACTIVITY					DOCUMENT						
	Objectives x2	Options x1	Execution x4	S+C x1	SCORE	Structure x0.25	Ortogr. x0.25	Gramm. x0.25	Format x0.25	Title x0.5	Filename x0.5	SCORE
(1.0) Excelent												
(0.8) Very Good												
(0.6) Good												
(0.4) Fair												
(0.2) Weak												
	2	0.8	3.6	0.7	7.1	0.2	0.25	0.25	0.25	0.5	0.2	1.65

cial. Recorreu-se essencialmente ao correio electrónico e numa fase mais final ao *instant messaging* para agilizar a comunicação. Devido a esta opção a capacidade de comunicação escrita foi muito desenvolvida.

3 SOCIAL WEB ENGINEERING

Esta secção descreve as actividades relacionadas com a realização de *Social Web Engineering* na página de Facebook da *HackerSchool*. Estas actividades resultaram na produção do relatório em anexo, que foi enviado à entidade promotora.

3.1 Análise das Impressões das Publicações

3.1.1 Como foi realizada

A análise, da responsabilidade do André Baltazar, das impressões das publicações na página da *HackerSchool* foi realizada através da utilização da Graph API do Facebook em conjunto com o Facebook SDK para PHP. O SDK foi utilizado para descarregar as últimas 250 publicações da página em formato JSON. Foi criado um script PHP para processar estes dados.

Foi calculada a média das percentagens (fãs e não fãs) para cada tempo e depois foi multiplicado pelo número de publicações em cada um desses intervalos de tempo e por fim divididos pelo máximo desse resultado. Por fim, foram produzidos gráficos com estes resultados. Foi feito também análise utilizando um método muito pouco diferente deste mas que apresenta resultados piores. Esta análise pode ser consultada no relatório enviado para a *HackerSchool* que está em anexo a este.

3.1.2 Principais Resultados

Os principais resultados obtidos através desta análise foram que as melhores horas para atingir todos os utilizadores são às 13 horas, 16 horas e das 21 horas as 22 horas. Os gráficos referentes a estes resultados estão na secção 3.2 do documento em anexo.

3.2 Análise de relação entre Likes e outros factores

Esta análise, da responsabilidade do Manuel Reis, teve como principal objectivo tentar relacionar o sucesso de uma publicação, medido de acordo com o número de likes da publicação com características da própria publicação, i.e., tamanho do texto, inclusão ou não de conteúdos multimédia, termos utilizados, e conteúdo emocional do texto.

3.2.1 Como foi realizada

O conjunto de posts publicados pela *HackerSchool* na sua página foram recolhidos com recurso a Graph API do Facebook. Posteriormente foram aplicadas algumas técnicas de Língua Natural e Análise Estatística para tentar relacionar as referidas características das publicações com o respectivo número de likes.

3.2.2 Principais Resultados

Concluiu-se que publicações com textos curtos, emocionalmente positivos e com inclusão de fotografias têm mais sucesso. Foram ainda apresentados quais os termos e temas que a *HackerSchool* deve abordar nas suas publicações para ter uma comunicação mais eficaz nas redes sociais.

3.3 Análise das Estatísticas Gerais

Esta análise, da responsabilidade do João Alexandre, teve como objetivo fazer uma observação mais geral das estatísticas obtidas. Visa observar a evolução dos gostos e do alcance das publicações na página do Facebook da *HackerSchool*, assim como observar as características do público-alvo e ainda ver a relação entre os gostos, comentários e o alcance com as características das publicações.

3.3.1 Como foi realizada

As análises foram feitas através das permissões dadas pela *HackerSchool*, na página do Facebook, e foram extraídas através das observações feitas aos gráficos fornecidos pelo Facebook assim como pela Graph API, que permitiu retirar as conclusões relativamente às publicações. Os dados retirados foram organizados para conseguir uma melhor perspetiva em relação aos objetivos visados.

3.3.2 Principais resultados

Através das análises feitas, concluiu-se que o público alvo está, maioritariamente, na faixa etária dos 18-24 anos e que são principalmente homens a visitar a página. Foi observado que os gostos têm uma evolução lenta devido também ao alcance das publicações ser baixo. Assim sendo e observando as publicações, concluiu-se que as publicações deveriam ter mais fotos e vídeos para que o alcance, os gostos e os comentários sejam mais elevados.



Manuel Reis Aluno de mestrado de Engenharia Informática de Computadores no Instituto Superior Técnico.

Área principal de estudos: Sistemas Distribuídos *Área secundária:* Sistemas de Informação Empresariais.

3.4 Resultados gerais

As análises realizadas tiveram como resultado uma série de recomendações a serem seguidas pela *HackerSchool* para tentar aumentar o nível de popularidade das publicações na sua página de Facebook. Estas recomendações são que os conteúdos textuais devem ter um tamanho reduzido e ser emocionalmente positivos, devem incluir fotografias e devem ser publicados entre as 21h e as 23h. Como resultado foi também sugerida uma lista de termos que devem ser usados nas publicações.



André Maurício Baltazar Aluno de mestrado de Engenharia Informática de Computadores no Instituto Superior Técnico.

Área principal de estudos: Engenharia de Software *Área secundária:* Sistemas Multimédia

4 CONCLUSÃO

Consideramos que as actividades descritas ao longo deste relatório, i.e., actividades inerentes à própria organização de um trabalho colectivo e actividades relacionadas com *Social Web Engineering* foram extremamente relevantes para a nossa formação enquanto futuros engenheiros. Estas actividades fizeram-nos abordar questões humanísticas e resolver problemas não técnicos com os quais normalmente não nos deparamos no decurso de um currículo académico na área da engenharia. Isto leva a que, a nosso ver, esta actividade esteja totalmente de acordo com o espírito da disciplina de Portefólio.

Neste tipo de documento (Técnico)
a Conclusão deve começar com
um resumo do assunto abordado
e depois deve falar o resultado



João Alexandre Aluno de mestrado de Engenharia Informática de Computadores no Instituto Superior Técnico.

Área principal de estudos: Engenharia de Software. *Área secundária:* Sistemas Distribuídos.

AGRADECIMENTOS

Os autores agradecem a Rui Santos Cruz, Paulo Carreira, e Inês Davim pelo apoio demonstrado no desenrolar da actividade.

APÊNDICE

RELATÓRIO HACKERSCHOOL



Relatório de Social Web Engineering

Hackerschool: Facebook

André Baltazar, João Alexandre e Manuel Reis

Dezembro 2014

1 Introdução

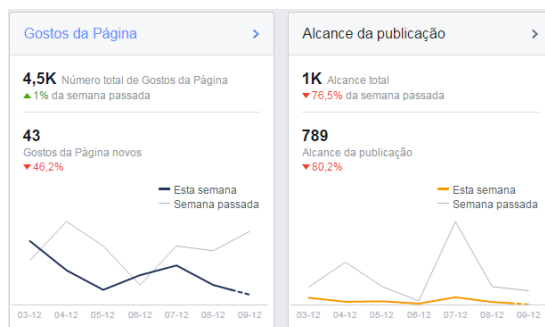
Este trabalho surgiu no contexto da execução de uma actividade extracurricular exigida pela disciplina de Portfólio Pessoal do curso de Engenharia Informática de Computadores.

O principal objectivo deste relatório é apresentar um conjunto de recomendações que os editores da página de Facebook da HackerSchool possam seguir com o intuito de melhorar a popularidade desta mesma página. Com este fim em vista apresentamos um conjunto variados de análises, que vão desde análise de estatística gerais até à tentativa de correlacionar posts populares com determinadas características desses mesmos posts.

2 Análise das Estatísticas Gerais

2.1 Vista Geral

Pelo observado repara-se que o nº de gostos totais da página tem aumentado muito lentamente. Tendo aumentado 1% dos gostos em relação à semana passada (1/12 - 7/12). O aumento do nº de gostos da semana passada (1/12 - 7/12) em relação à semana anterior teve uma baixa de 46,2%, o que indica que o alcance das publicações e gostos na mesma da página têm tido pouco alcance, assim como a própria página. Tal como tinha dito o alcance das publicações teve uma baixa em relação à semana passada (1/12 - 7/12) de 80,2%.



(a) Overview dos gostos e alcance da página HackerSchool nesta semana.

2.2 Gostos

O nº de gostos no último mês (Novembro) aumento de 4.311 para 4.529, este aumento de 218 gostos dá uma média de 7 Gostos/Dia. Repara-se também que tem existido um baixo cancelamento de gostos sendo no máximo 2 gostos por dia.

Publicações com vídeos e fotos têm mais gostos e comentários que publicações com simples ligações.

2.3 Alcance das publicações

O alcance varia consoante as publicações que aquele certo dia tem e pelo observado no gráfico seguinte, dá para verificar que as publicações em que existem vídeos têm um maior alcance que as publicações com fotos ou com ligações apenas.



(a) Alcance das publicações no último mês.

2.4 Pessoas que gostam da página

Em termos de géneros, apenas 18% dos fãs são mulheres enquanto existem 81% de homens. Em termos de idade, a maioria dos fãs da página, como é normal, rondam os 18-24 anos (63%).

2.5 Conclusões Gerais

Dado o género e idades dos fãs e também o alcance que certas publicações tiveram achamos que as publicações deveriam ser feitas depois do jantar entre a 21h até às 23h para atingir mais alcance e ter mais gostos e partilhas.

Concluimos também que deverão dar mais ênfase às publicações com imagens e vídeos visto que artigos e ligações têm pouco alcance. Damos um exemplo, a Apresentação do Shark Tank teve um alcance bastante baixo para o quão interessante o assunto é, no nosso entender isto deve-se porque as publicações apenas têm o próprio evento. Talvez umas imagens acerca do que é o Shark Tank ou um vídeo (feito por vocês ou um trailer do Shark Tank americano) tivesse tido mais impacto, além disso as publicações foram feitas num horário ao qual consideramos errado como já referimos no ponto acima.

3 Análise do Alcance das Publicações

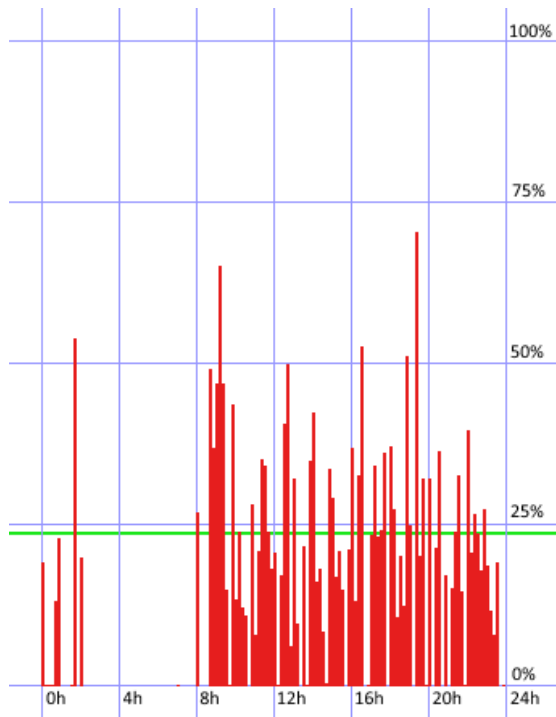
3.1 Análise de impressões de publicações para não fãs (método 1)

Para aumentar a exposição da HackerSchool é interessante olhar para a percentagem de pessoas não fãs que a publicação atinge e ver a que horas essa percentagem é maior.

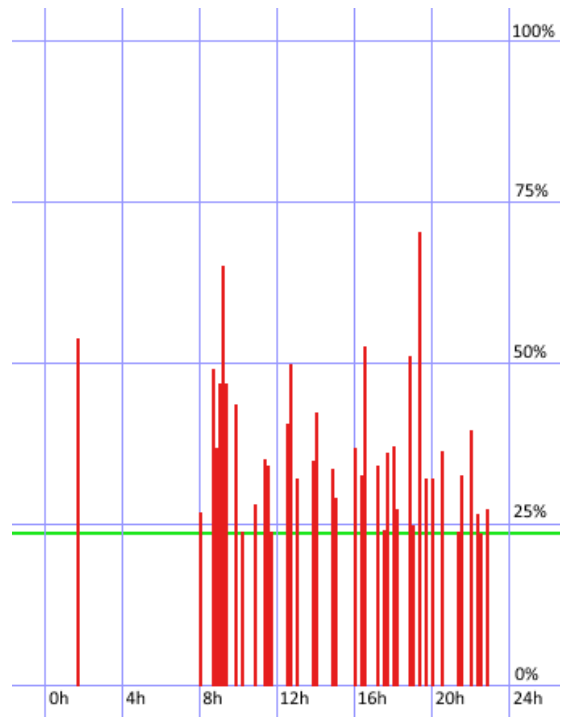
O gráfico 3a apresenta essas percentagens para as ultimas 250 publicações (até 30 Novembro) na página da HackerSchool. Reduzindo o gráfico para apenas periodos de 10 minutos que estão acima da média de publicações (linha verde em todos os gráficos), representado pelo gráfico 3b, não conseguimos ainda facilmente identificar as horas em que a exposição da publicação é maior.

Para tentar descobrir então um periodo mais pequeno em que os posts têm mais exposição, podemos olhar para o número de histórias criadas(apresentadas nas Feeds de Noticias, etc..) e tomar apenas os periodos em que estão abaixo da media da percentagem de histórias criadas, pois estes são os posts que tiveram mais influencia. O resultado desta filtragem é o gráfico 3c que permite distinguir claramente 2 ou 3 periodos que passam nestas condições. Podemos considerar ainda que posts que encontram 5% acima da média da percentagem de histórias criadas ainda são relevantes e com esse critério é foi gerado o gráfico 3d. Este apresenta quase os mesmos periodos que o anterior, sendo estes aproximadamente das 11:00 às 12:45, das 17:30 às 20:00 e por fim das 21:30 às 22:30.

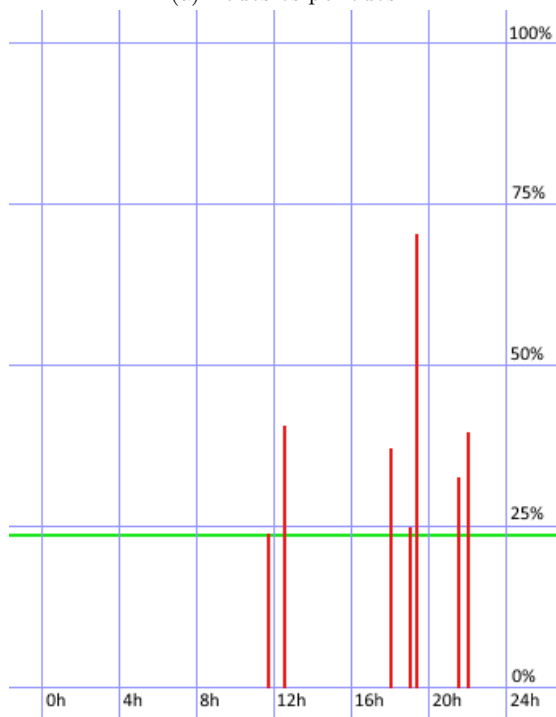
Esta análise apresenta os valores que tiveram mais sucesso para atingir pessoas não fãs da página, sendo que a publicação durante esses períodos não garante uma maior exposição, no entanto são os que apresentam maior possibilidade da mesma.



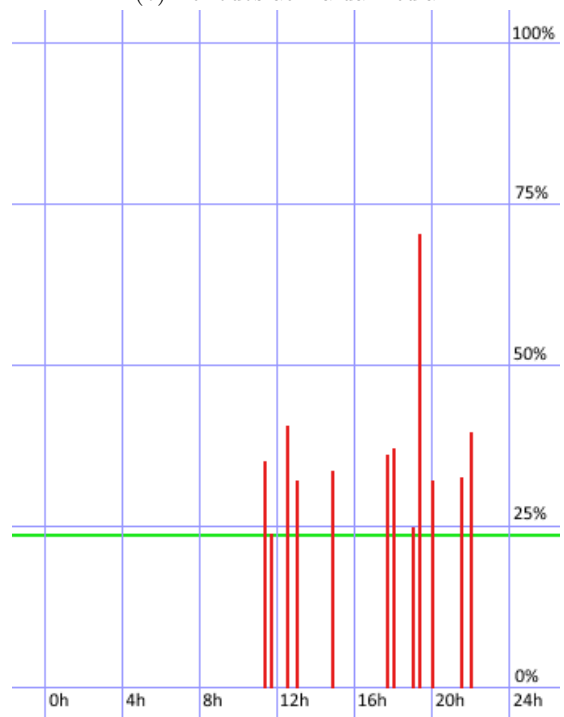
(a) Todos os períodos



(b) Períodos acima da média



(c) Períodos acima da média com percentagem de histórias criadas abaixo da média



(d) Períodos acima da média com percentagem de histórias criadas abaixo da média mais um desvio de 5%

Figura 3: Percentagem média de não fãs que visualizaram a publicação (dividido em períodos de 10 minutos).

3.2 Análise de impressões de publicações para fãs e não fãs (método 2)

Esta análise tem o mesmo objetivo que a da secção 3.1, mas neste caso o método é diferente pois este tem em conta o número de publicações que ocorreu nesse período de tempo, não sempre apenas uma média simples. Foram usados os mesmos dados que no método 1, mas não foi considerado o número de histórias criadas, apenas os números de impressão das publicações.

Para representar estes gráficos foi primeiro calculado a média das percentagens para cada período de tempo e depois multiplicado pelo número de publicações de cada um desses períodos e por fim foram divididos pelo máximo desses resultados para que o gráfico final ficasse entre 0 e 1.

Esta análise devolve resultados muito semelhantes aos que foram apresentados para o método 1, mas ainda assim apresenta, algumas diferenças, por exemplo no gráfico 4b podemos ver que existe algum sucesso nas publicações perto das 9 horas da manhã.

Analisando ambos os gráficos da figura 4 pode-mos ver que as melhores horas para atingir todos os utilizadores poderá ser perto das 13 horas, 16 horas ou das 21 às 22 horas.

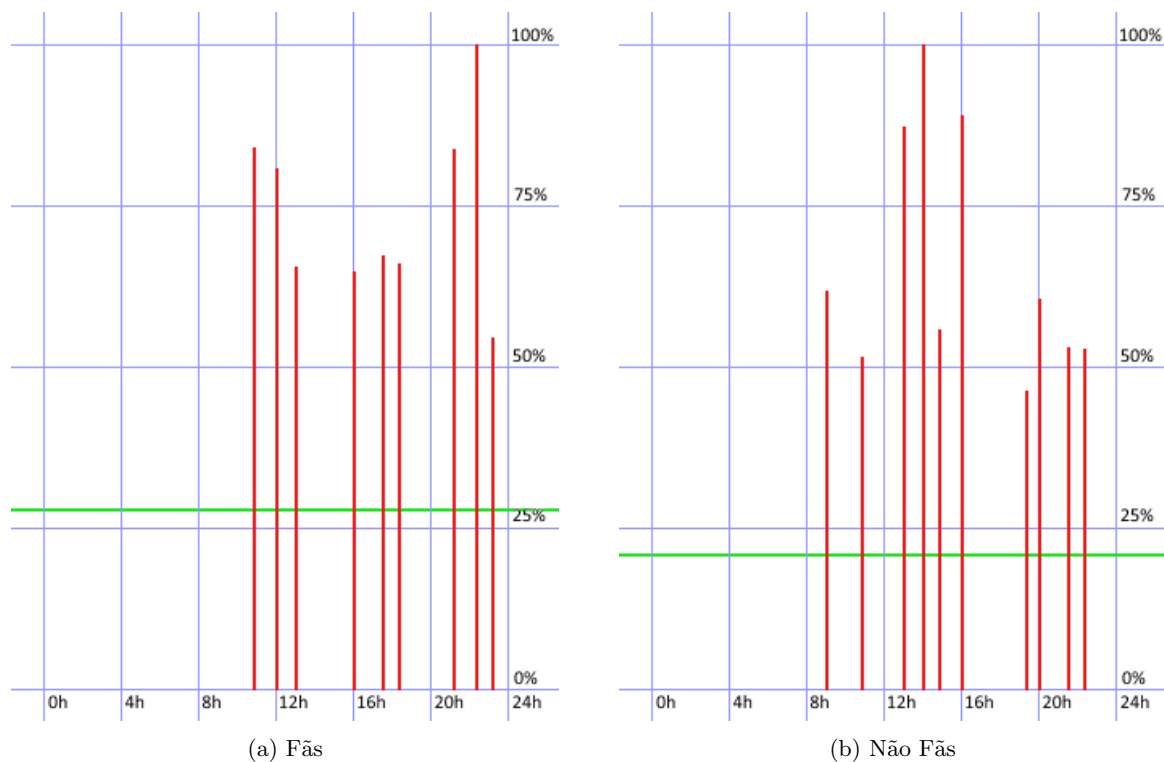


Figura 4: Impressões das publicações - Método 2

4 Análise da Relação entre *Likes* e outros factores

Os dados obtidos para sustentar a análise apresentada nesta Secção foram obtidos através do método *feed* da *Facebook Graph API*¹. Conseguimos assim recolher todos os posts publicados pela HS no seu próprio feed. Para cada um destes posts foi considerado: (1) o número de likes, (2) o conteúdo textual, e (3) informação sobre conteúdo multimédia. Nesta secção, utilizamos como métrica de sucesso para um dado post, o seu número de likes. Foram considerados 1074 posts (a totalidade dos disponíveis na data da extracção).

¹<https://developers.facebook.com/docs/graph-api/reference/v2.2/page/feed#read>

4.1 Likes v.s. Conteúdos fotográficos

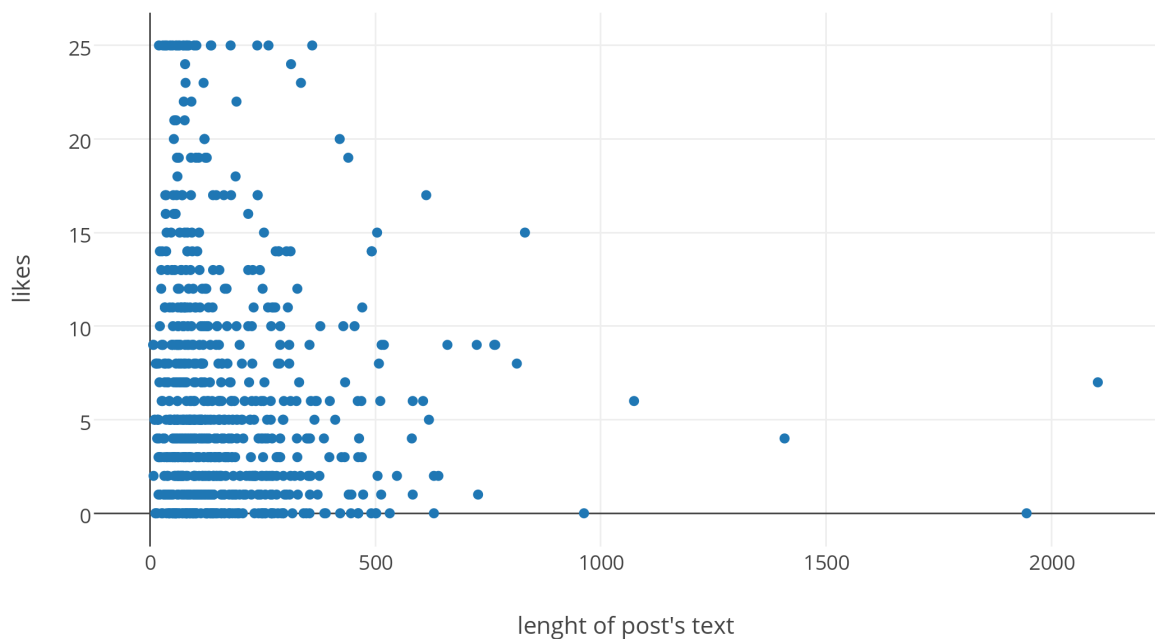
Esta Subsecção apresenta a relação entre a existência ou não de conteúdos fotográficos com o número de likes obtidos. Analisando as estatísticas apresentadas na Tabela 1, podemos observar uma clara relação entre um maior número de likes e a inclusão de fotografias nos posts. Um post sem fotografia tem em média 1.99 likes enquanto que um post com conteúdo fotográfico tem em média 7.75 likes.

	Posts	Likes	Média de likes por post
Com Foto	532	4124	7.75
Sem Foto	542	1081	1.99

Tabela 1: Relação entre conteúdo fotográfico num post e número de likes

4.2 Likes v.s. Comprimento textual dos Posts

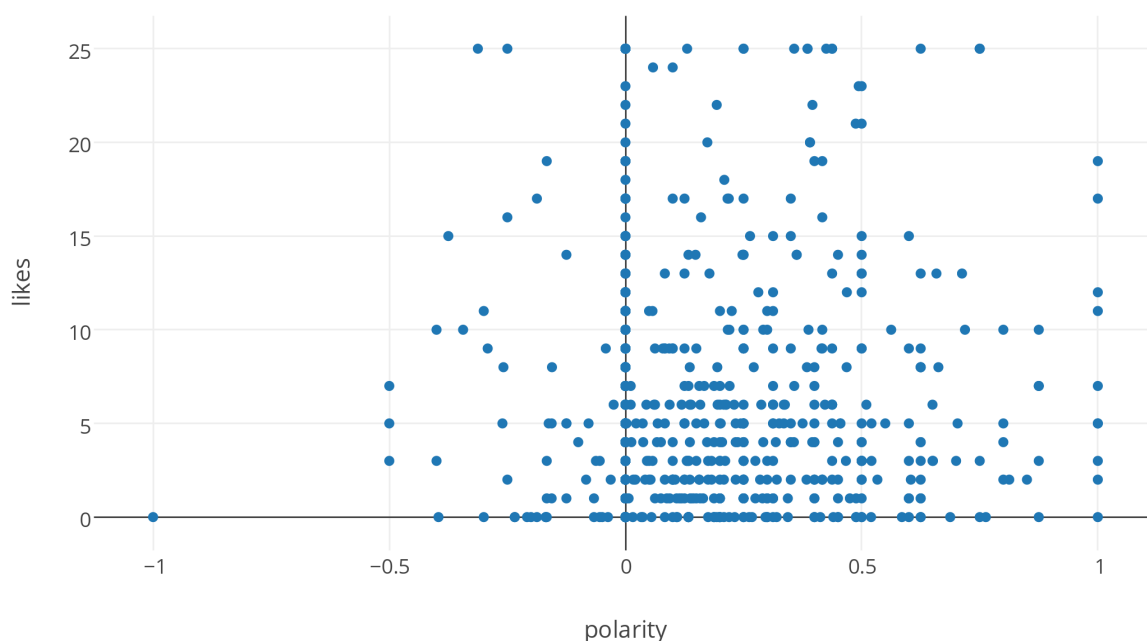
Esta Subsecção pretende estudar se existe relação entre o número de likes que um post tem e o respectivo comprimento do seu conteúdo textual. Observando o gráfico abaixo apresentado conseguimos observar uma clara relação de proporcionalidade inversa entre o comprimento do texto e o número de likes obtidos.



4.3 Likes v.s. Polaridade emocional das mensagens textuais

O conteúdo emocional de uma mensagem (i.e., uma mensagem com conteúdo textual positivo ou negativo) pode ter consequências quanto a popularidade da mesma. Esta Subsecção estuda essa relação. Com este propósito, sujeitamos cada um destes conteúdos textuais a uma avaliação da polaridade emocional do mesmo.

Esta avaliação retorna, para cada conteúdo, um valor entre -1 e 1, sendo -1 um conteúdo textual muito negativo, e 1 um conteúdo textual muito positivo. Esta avaliação da polaridade emocional foi realizada usando a biblioteca *Textblob*² (Todos os textos tiveram de ser previamente traduzidos para inglês antes do uso desta ferramenta). O gráfico abaixo apresenta a polaridade emocional de uma mensagem com o número de likes do post que contém essa mensagem. A sua interpretação não oferece conclusões muito consistentes, contudo podemos inferir, por exemplo, que os posts que têm um número elevado de likes (e.g, 25), são na sua maioria posts com mensagens positivas.



4.4 Likes v.s. Termos utilizados nos conteúdos textuais

Consideramos útil inferir quais os termos mais utilizados nos posts mais populares, pois assim conseguimos inferir quais são os temas que surtem mais interesse nos fãs da página de Facebook da HS. Para este efeito seleccionámos os 25% de posts com mais likes e removemos dos seus conteúdos textuais as *stopwords* (i.e., palavras sem significado ex: e, ou, mas) utilizando o corpus *NLTK*³ para português. A imagem, obtida com recurso ao *Wordle*⁴ abaixo apresentada, tenta apresentar uma visualização dos resultados obtidos. Nesta imagem, os termos que aparecem com maior dimensão são os mais usados nos posts mais populares. Alguns dos termos interessantes que se destacam mais são *workshop*, *lx*, *reactor*, *projectos*, *design*, *amanhã*, *semana*, *vai*, *dia*, *empreendedorismo*, *arduino*. A aparição do termo *http* pode sugerir o interesse por parte dos fãs em relação a posts que incluem links.

²<https://textblob.readthedocs.org/en/latest/#>

³<http://www.nltk.org/>

⁴www.wordle.net

