

Writing for the web: An editorial framework for the development of objective news headlines

By

Chris Blatchford

A DISSERTATION

Submitted to

The University of Liverpool

in partial fulfilment of the requirements
for the degree of

MASTER OF SCIENCE

28/08/2014

ABSTRACT

Writing for the web: An editorial framework for the development of objective news headlines

By

Chris Blatchford

Much has been written on how to write for the web, and in some cases this has extended to news headline development, but for the most part literature has focused on generic best practice writing principles. These principles can be applied to digital news headline development, along with additional subjective factors, which influence the structure of the text; these influential elements are manifested through structural elements such as content type, layout and format, but are on the whole applied randomly. This paper investigates the assumption that these elements can be identified, ranked and ordered in a particular sequence which, when applied to a news headline or story, results in the most objective, consistent and factually accurate news headline for the associated news content – it is also the assumption of this paper that these ‘best practice’ headlines are the most preferred by end users. To test this assumption, an interview was conducted with Thomson Reuters employees, results of which are presented in this paper, and used to build a best practice framework. A subsequent survey of Thomson Reuters employees was conducted to test the best practice framework, through generating news headline options for end users to indicate a preference. Results show that certain best practice elements are more influential on end users than others, certain combinations of elements are more effective than others, and that the content of the news story (financial versus scientific in this case) can influence end user preference, and therefore which element should be considered ‘high priority’.

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

Chris Blatchford

“This dissertation contains material that is confidential and/or commercially sensitive. It is included here on the understanding that this will not be revealed to any person not involved in the assessment process.”

Student, Supervisors and Classes:

Student name:	Christopher Blatchford
Student ID number:	H00019894
GDI name:	Bendick Makleko
CRMT class ID:	Comp Research Methods Training LAUR-906-201424-2
DA name:	Paul Kingsley
CAC class ID:	Computing Advisor Class CKIT-702-201424-8

ACKNOWLEDGEMENTS

I would like to thank my dissertation advisor Paul Kingsley for his help, guidance and support throughout - his constructive and timely feedback helped me finalise and submit this paper whilst becoming a father (time became a precious commodity). I would like to thank Thomson Reuters, for allowing me to perform interviews and distribute surveys to employees, and a special thanks to Mark, for the time he spent speaking with me around this subject. Finally, I would like to thank my wife Carley, for just being there when I needed her.

TABLE OF CONTENTS

	Page
LIST OF TABLES	3
LIST OF FIGURES	4
Chapter 1. Introduction	5
1.1 Scope	5
1.2 Problem Statement	6
1.3 Approach	7
1.4 Outcome	9
Chapter 2. Background and review of Literature	10
2.1 Background	10
2.1.1 Writing for the Web	10
2.1.2 Digitizing news	12
2.1.3 News headline structure	13
2.1.4 Cognitive and contextual considerations	14
2.1.5 Editorial process.....	16
2.1.6 Text summarization, and Natural Language Processing (NLP)	18
2.2 Literature Review	18
Chapter 3. Theory	20
3.1 Hypotheses	20
3.1.1 Headline Content (highest to lowest)	21
3.1.2 Headline Structure (highest to lowest)	21
3.1.3 Headline Structure (highest to lowest)	21
Chapter 4. Analysis and Design	24
4.1 Interviews Design	24
4.1.1 Description, basis & justification	24
4.1.2 Interview preparation	25
4.1.3 Interview Participant selection	26
4.2 Survey Design	26
4.2.1 Description, basis & justification	28
4.2.2 Survey Participant selection	28
4.2.3 Use of priority ranking items in workflow	28
4.3 Framework	29
4.3.1 How the framework will be built	29
Chapter 5. Methods and Realization	30
5.1 Approach	30
5.2 Interview Results	30
5.2.1 Readership Focus	30
5.2.2 Thomson Reuters ‘Do’s and Don’ts’	31

5.2.3	Editorial Authorship.....	34
5.3	Interview output and Utilisation	35
5.4	Survey	35
5.4.1	Survey Implementation.....	35
5.4.2	Survey results analysis method.....	36
5.5	Framework Design	36
5.5.1	The framework design.....	36
Chapter 6.	Results and Evaluation	40
6.1	Survey Results	40
6.2	News headline criteria, a refresh	40
6.3	Results overview, by hypotheses	41
6.4	Overall analysis, financial and scientific	44
6.5	Overall analysis, financial versus scientific	47
6.6	Overall analysis, format	49
6.7	Open ended question responses	49
6.8	Hypotheses testing summary	52
Chapter 7.	Conclusions	54
7.1	Lessons Learned	54
7.2	Academic Application and Limitations	54
7.3	Business Application and Limitations	54
7.4	Recommendations for Future Research & Work	55
	REFERENCES CITED	56

LIST OF TABLES

	Page
Table 1 Interview Methods	25
Table 2 Content Test 1	41
Table 3 Content Test 2	42
Table 4 Content Test 3	42
Table 5 Structure Test 1	43
Table 6 Structure Test 2	43
Table 7 Structure Test 3	44
Table 8 Format Test 1	44
Table 9 Overall Analysis Financial & Scientific	45
Table 10 Overall Analysis Financial & Scientific, Actual	45
Table 11 Overall Analysis Financial & Scientific, Hypotheses	45
Table 12 Original Headline Preference Analysis.....	46
Table 13 Overall Analysis Financial Versus Scientific	47
Table 14 Overall Analysis Financial & Scientific, Actual	47
Table 15 Overall Analysis Financial & Scientific, Hypotheses	47
Table 16 Overall Analysis Financial.....	48
Table 17 Overall Analysis Financial, Actual	48
Table 18 Overall Analysis Financial, Hypotheses	48
Table 19 Overall Analysis Format	49

LIST OF FIGURES

	Page
Figure 1 Research Phases.....	8
Figure 2 Survey Question	27
Figure 3 News Readership Categories	31
Figure 4 Interview Headline Recommendations.....	32
Figure 5 Survey Question	36
Figure 6 Framework Design	37
Figure 7 Framework Content Design.....	37
Figure 8 Framework Structure Design.....	38
Figure 9 Framework Format Design.....	38
Figure 10 Survey Responses	40

Chapter 1. INTRODUCTION

1.1 Scope

The aim of the research is to understand how online news headlines are interpreted and digested by end users, taking into account content, structure and formatting. Through an understanding of these elements, we will gain further insight into how news editorial processes can further improve and streamline their authorship production operations, whilst remaining objective and eliminating any subjective elements from the end news headline.

The primary proposed solution is the development of a best practice framework, which can be used to analyze news headline content, structure and presentation. A secondary deliverable (out of scope currently) of the research would include a software processing application, which through a text analytics algorithm based upon the primary framework deliverable, provides recommended headlines based upon content, structure and formatting.

The document will be split into three main areas; research of the literature and existing industry practice, development of a best practice framework, and the subsequent experiment and evaluation phases. The best practice framework will be driven by my initial hypotheses around best practice news headline development – both the framework and the experiment output will test these hypotheses validity. The primary deliverable within the document is the framework itself, which based upon the hypotheses stated, will confirm or reject the suggested approach.

The evaluation phase falls into two distinct parts; firstly, interviews with Thomson Reuters editorial staff, along with a review of existing TR best practice, to provide practical, field data on news headline structure. This in combination with the academic research will output the best-practice framework IT artefact.

The experiment will consist of running each headline through the framework process, making recommendations on content, structure and format. This outputs a number of best practice produced headline data (experimental) for analysis and review. These are then paired with the externally collected data (control) for comparison and ranking purposes. Results will be evaluated by internal Thomson Reuters employees via a distributed/online survey.

1.2 Problem Statement

News Headlines are traditionally developed qualitatively via editorial process, making it a largely subjective process, given the variations in editorial approach and agenda – truly objective news headlines (and stories) must be devoid of political or social agendas, or any other influence, and deliver purely factual information on the news subject area. Outside of subjective influences, the quality of news headlines vary considerably depending on the skills and capability of the editorial function in control – quality in this case is measured in terms of content, structure and format.

Consistency and objectivity for any editorial team has always been a challenge, ever since William Caxton set-up his first print press in Bruges right through to the advent of the internet. Today, editorial teams have to also contend with the challenges associated with Big Data, and the vast amounts of information we're presented with on a daily basis. How quickly an organization delivers its news to consumers is absolutely critical, hence speed and automation are two of the most important elements of news delivery. Speed can be achieved through technical advancements, such as database efficiency and the proficient corralling of multiple news sources, similarly, automation can be achieved through the use of complex algorithms. However, prior to building an algorithm which defines and delivers a news story, editorial teams need to be confident the underlying mechanics of how the code structures the news headline is equal to or better than human editorial input, and ensure the delivered headline achieves high web rankings, strong click-through rates whilst maximizing cognitive effectiveness for the reader.

One of the ways in which editorial teams can achieve this automation is through the application of business process automation. Essentially a subset of process engineering, whereby all elements of a given process are identified and documented, automating that process is only beneficial once best practice is achieved; automating an inefficient or otherwise broken process achieves very little. In the specific example of news story generation, before an editorial team can start to think about automating that process, a best practice or framework needs to be defined and proven.

This brings us to the ultimate objective of this research - I will be proposing a best practice framework which could be used to facilitate the news headline editorial process, a methodology which editorial teams can utilize in order to provide a consistent, objective & informational technique of developing news headlines – the removal of subjective, influencing factors, to provide a factually efficient news headline.

1.3 Approach

Initially, an understanding of how to write for the web, how to author news headlines and how publishing of news headlines function operationally will be required in order to extract a best practice framework. The approach will be based upon pre-existing research and practical field data capture. The core data required for this purpose will be:

1. Thomson Reuters/external 3rd party & industry best practice editorial approaches.
2. Collection of news headline data (Thomson Reuters/external 3rd party & industry)
3. Experiment output (evaluation of framework produced headlines)

In order to achieve item number 1, an understanding of editorial process and procedure is required (headline generation & deployment) along with a high level comprehension of news writing best practice.

The research will utilise both **Review** and **Descriptive** design methods; Review will provide the initial literature review resulting in the suggested framework, with the Descriptive approach providing the structure for news headline case studies and surveys. For the

purpose of the surveys, we will be utilising an online survey tool, SurveyMonkey, to collect responses.

The project will run through four main stages of development:

1. Firstly, a review of existing literature and research papers on web authoring (including traditional literary reviews of writing best practice), understanding how the web/internet changes this dynamic and how writing for the web has transitioned over the past decade, resulting in our initial hypotheses.
2. Secondly practical research into Thomson Reuters scientific and financial news delivery, with interviews and questionnaires designed to understand what structure and delivery mechanism are the most favoured, whether existing methodologies and frameworks are utilised, and what the most effective mechanisms are.
3. Thirdly, based upon the research results, I will develop a framework which will provide recommendations on how to approach the subject matter.
4. Fourthly, the framework will then be evaluated utilising real world operational data in order to fully test and validate the hypotheses.

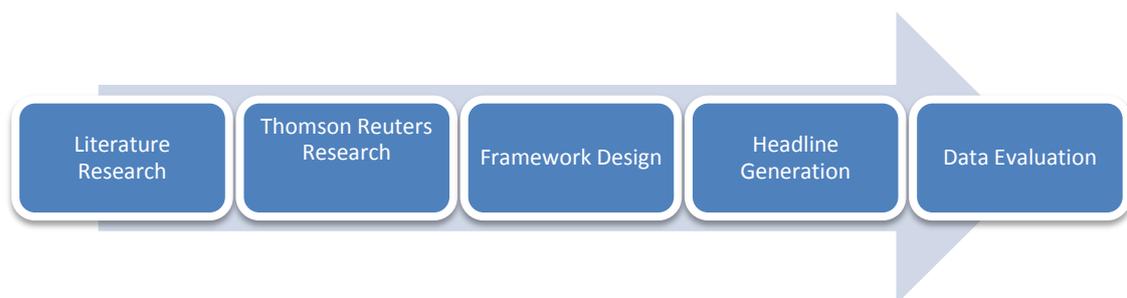


Figure 1 Research Phases

1.4 Outcome

The results of the interviews are utilized and fed into the best practice framework, along with extracted elements from the literature review. The best practice framework is then presented and explained. The results of the survey are presented and analyzed. Limitations of the survey and results are explored, areas of further research are presented and potential commercial applications are suggested.

Chapter 2. BACKGROUND AND REVIEW OF LITERATURE

2.1 Background

The main themes running through the researched literature can be summarised under four main headings (for the purposes of this research paper); writing for the web (generic), digitizing news, effective news headline authorship and cognitive and contextual considerations. These three themes serve to frame the more quantitative structure of headline content. For example, Morke & Nielsen (1998) suggests that “conciseness, scannability, and objectivity” are key tenants of writing for the web, resulting in “faster, more efficient processing of information”. This is part of a more general “rule set” that can be applied to authoring content online; however news content changes that paradigm once again. Dor (2003) goes into some detail around news headline authorship, defining newspaper headlines as “relevance optimizers: They are designed to optimize the relevance of their stories for their readers.”, which is a component part of the news headline structure, in addition to delivery of specific information in the minimal amount of space (especially for mobile device usage). Develotte and Rechniewski (2013) go into further detail around cultural and more detailed linguistic factors which may influence structure and interpretation; for example, they discuss perspective and the “role played by headlines in orienting the reader's interpretation of subsequent 'facts' contained in the article”. This perhaps speaks more to the more qualitative elements, such as cognitive processing and contextual knowledge, which have an indirect impact upon the readers' interpretation of the information contained within the headline. We also look at the editorial process, and how this affects news headline output – it is often the case that different editorial teams will output significantly different headlines based upon the same intelligence, indicating a particular social, political or other biased influence.

2.1.1 Writing for the Web

Morkes and Nielsen (1998) found through their earlier studies on ‘How to Write for the Web’ that the ability to scan text that is concise and objective, was favoured over ‘marketese’ (overly hyped promotional writing), and subsequently found that application of

these three rules led to a 124% usability increase over the original version (to a test website). Morkes and Nielsen (1998) basic premise is that conciseness, scannability and objectivity “reduces the users cognitive load, which results in faster, more efficient processing of information”. This is, and continues to be, a common theme amongst information interpretation, especially in 2013 & 2014 where the concept of Big Data is beginning to manifest itself in practical implementations; one such area is news ingestion, and how editorial teams can best structure their content to ensure maximum consumption, when faced with huge amounts of data from multiple sources. Morkes and Nielsen (1998) repeat their case study with a new subject area (Sun Microsystems technical whitepapers versus their original tourist information text) and, through application of their original guidelines in addition to a few additional rules (such as the use of Hypertext to “split long text into smaller and more focused pages”) were able to increase overall usability by 159%. Redish (2012) echoes Morkes and Nielsen’s thoughts, especially around conciseness and objectivity (albeit more informally) with recommendations around both length of headline (conciseness) and the use of a “statement, question, or call to action” (objectivity). Redish also recognises the need for conciseness in headlines, to promote scannability and overall ingestion & interpretation time (processing and storing of information & subsequent utilisation (I cover news headline development more fully in section 1.2.3). For example, consider the following news headline and excerpt:

Sea creature fossil found with oldest-known cardiovascular system

(Reuters) - You've heard of having a heart of stone, but this old guy takes it literally. Scientists said on Monday they have found a fossil of a shrimp-like creature that lived 520 million years ago with an exquisitely preserved heart and blood vessels that represent the oldest-known cardiovascular system.

This headline contains key information, is concise, and if scanning a website one will register several key elements such as ‘fossil’ and ‘cardiovascular’, which should be sufficient to warrant further detailed reading should the headline be of interest or relevance to the reader.

Building upon the research around scannability, when viewing content from a format perspective as opposed to through a content/structural lens, the general consensus around format appears to be that font style has a significant impact on information consumption. Bernard and Lida (2002), in their study of online font preference, focused around eight popular fonts, and came to several conclusions around best practice for font style and format (speed of reading, legibility and 'attractiveness'), leading to an overall recommendation that the font face 'Verdana' was the most preferred out of the evaluated types.

2.1.2 Digitizing news

Along with the basic principles for writing for the web, the digitization of news content has undergone a significant amount of change and development over the last decade. Not only do we need to consider conciseness, scannability and objectivity, we need to also consider the device being utilised to view the content, the content's relationship to other content (online, including SEO) and the subject matter in question. As Caumont (2013) outlines, based upon research from the Pew Research Center (2013 State of the News Media report) 19% of Americans "saw news on a social network "yesterday" in 2012", which given the limited space and text characters available on networks such as Facebook or Twitter, changes the way in which we present and ingest news content. In addition, Caumont found that 64% of tablet owners and 62% of Smartphone owners "said they got news on their devices in 2012", which again changes the delivery mechanism, and thus the development and authorship approach to content. The internet also provides a new type of news structure sitting behind the headline; the concept of longevity, or how long content remains online, has a significant impact on how headlines are interpreted. Chua (2014) suggests that the "fast" (headline) driven content should provide the building blocks to the more "in-depth" content, that the two content sets should "speak" to each other in some manner. This again potentially changes the content of a news headline, with assumed knowledge from a wider, persistent lake of knowledge held on the internet. Lastly, it is important to recognise the evolution of news headlines, and news content in general, from the original ink based paper format through to today's digitally enhanced,

hyperlink enabled content. Boczkowski (2013) suggests that “cultural and material changes do not proceed in a historical vacuum, but are influenced by the legacy of processes that preceded them”; this is the primary rationale for reviewing both writing for the web concepts along with how to structure news headline content, both factors of which are strong influencers of today’s online news headline structure. Gregory (2012) also discusses the suggested differences between ink and digital, citing 7 popular guidelines for writing news content, but conversely provides an opposing view, suggesting that in fact many of the web writing guidelines can (and should) be applied to ink based news content. Gregory suggests that in fact we are focusing on differences perhaps too much, and that “many of the fundamental writing issues that communicators should consider appear to apply in both print and web environments”.

2.1.3 News headline structure

Much has already been written around how to structure news headlines to increase end user consumption rates (essentially the number of users “clicking” through the headline to the main story content); Halbrooks (2014) suggests that a “focus on descriptive words and phrases” provides the most impact per word, Rogers (2014) states that being accurate, direct and use of the “subject-verb-object” ‘formula’ will increase click through rate, and Edelstyn (2014) goes as far as to suggest that “eight is the magic number” (of words in each headline), including a colon or subtitle and structuring the headline as a question all increase overall click through rate. Looking back at our example headline ‘Sea creature fossil found with oldest-known cardiovascular system’ we see the use of the ‘subject-verb-object’ structure - the subject is the ‘sea creature fossil’, the verb is ‘found’ and the object the ‘cardiovascular system’; the headline also adheres to Edelstyn (2014) eight word rule (if we consider oldest-known as a single word phrase).

However structuring headlines according to the audience, and understanding how the structure impacts interpretation and processing, is less well understood. Dor (2003) study into newspaper headlines as relevance optimizers (variables that increase the overall relevance of the news story, considering the audience and context), recognises the need

to understand readers – “their state-of-knowledge, their beliefs and expectations and their cognitive styles”, which can be somewhat extrapolated from the type of news being produced. Dor (2003) continues, and functionally defines newspaper headlines as “relevance optimizers: They are designed to optimize the relevance of their stories for their readers.”. Dor expands upon this concept by explaining that the headline also reduces the overall processing effort by the reader, whilst still ensuring a sufficient amount of “contextual effects are deducible within the most appropriate context possible”, so for example, with a financially driven news headline about the current state of Company X stock, the headline needs to provide relevant information, with some assumed knowledge of background information (contextual, presupposed between source and reader) whilst maximizing process efficiency; this enables the reader to ingest and use the information to make effective decisions quickly. Looking at our scientific news headline, we can see that this headline omits the ‘520 million year’ statistic, but the presence of ‘fossil’ implies an ancient artefact – the use of ‘fossil’ is sufficient, we don’t need to include the actual statistic. Dor, through his research, was able to extract 10 properties for defining the ‘appropriate headline’ including “headlines should be as short as possible”, “headlines should be clear, easy to understand, and unambiguous”, “headlines should be interesting” and “headlines should contain new information”.

2.1.4 Cognitive and contextual considerations

News headlines are by their nature subjective, vulnerable to a particular political, social or other influencing factor. We can see this in the popular UK tabloids and broadsheets; for example, the “Sun” typically leans towards the UK political Labour Party, whereas the “Daily Mail” in recent years has been far more right wing, supporting the Conservative Party. Then there are the extreme supporters, such as the “Daily Mirror” (Labour) and the Daily Telegraph (avowedly Conservative). The issue of course is that headlines are then edited and influenced by both the company and readership alike, and therefore structured accordingly, to impart a message that is far from impartial.

Dor (2003) touches upon this issue, noting that newspaper headlines don't always summarize the related stories; he states that "some headlines highlight a single detail extracted out of the story, and others contain a quotation which the editor decided should be promoted to the foreground". This is a good example of an external subjective factor to the story "data" influencing the structure and interpretation of the wider story. Taking our example headline 'Sea creature fossil found with oldest-known cardiovascular system', the headline omits details such as the age of the fossil, '520 million years', and the description 'shrimp-like creature', along with the statement 'exquisitely preserved heart and blood vessels'. The focus of the headline is clearly the cardiovascular discovery, which speaks to the scientific source the headline was published from, however an historian or archaeologist may find the fossil more interesting, and the way in which it has been 'exquisitely' preserved. Subjective headlines then can lead the reader, often into established clichés and prejudice, triggering frames and belief systems in the readers mind that are subsequently resolved in the ensuing text (Lindemann, 1990). Even if the choice of key words or phrases was not as a result of some political or other motivation, the entities selected by an editor have a direct impact upon interpretation. Zhang (2013) found that "global news headlines involve working with discourse that is heavily mediated and recontextualized, in which the transeditors put their own knowledge and values into the transedited texts.", which is essentially speaking to influencing factors that we recognise above. However, Zhang continues to suggest that "an absence of such discourse in news headlines, or even an absence of reporting on a particular event, may also be an indication of a stance adopted by a news agency."; editorial process that abstains from transediting could in itself suggest a particular stance, hence it is this qualitative, editorial process which is critical to news story and headline development.

Informational news headlines then should serve a number of purposes; at the very least, they should present (impartial) facts about the underlying story, they should contain relevant text to the story in question, and they should be simple to interpret and understand. Matheson (2004) suggests that the epistemology of journalism - the way in which journalism operates a knowledge-producing practice, has a direct impact upon how news is re-

ported, specifically that "the conventions of newswriting do not simply chronicle the world but that they constitute certain claims to knowledge about such matters", both from the journalist and audience perspective, and how the relationship between the two influences the news story. This speaks once again to the editor's job of selecting the 'right' information to be used within the headline, given the audience and journalistic source in question. Dor also picks up on this relationship, through reference to Iarovici and Amel (1989), who state that a news headline has a double function; semantic, regarding the referential text, and a pragmatic function, regarding the reader. They continue by concluding that the "main function of the headline ... is to alert the reader (receiver) to the nature of the content of the text". Thus, headline content, structure and format all combine to achieve a best practice for structuring the headline, with editorial processes further augmenting the headline, taking into account journalistic source and audience influencing factors.

2.1.5 Editorial process

There is also an indirect influence on the headline through the story itself, in the manner the wider network of internet & digital news will influence the story text via 3rd party sources. Hyperlinks to related stories may well provide alternative views, or different information entirely; Matheson (2004) states that weblogs (essentially online journalism) "produces meaning also within a notion of knowledge as a process and matter of connection, rather than contained within one text", hence the story is not a standalone piece of work as with traditional newspapers. This would imply the headline itself could draw details from outside of the realm of the related story, resulting in a mixed message, which is why understanding the objective of the headline, and the end user audience, is crucial. Understanding the news audience is then of great importance; whether they have prior knowledge of headline entities or related stories, or presupposed assumptions or expectations around the subject matter; essentially, how relevant a story is to an individual. Sperber and Wilson (2004) developed 'Relevance theory' and 'relevance optimization', in an "attempt to reduce a very complex set of phenomena having to do with communication and interpretation to a very constrained set of explanatory, cognitive notions", or "cognitive cost-effectiveness". Relevance theory is essentially how relevant a news headline is

to an individual, taking into account the number of contextual effects i.e. drawing upon a subset of existing assumptions about the individual words described within a headline. Thus, when looking to structure a headline, consideration must be given for the likely audience and their associated knowledge (and assumptions) around the words used. Looking back at our example news headline 'Sea creature fossil found with oldest-known cardiovascular system', the decision to exclude the term 'Scientist' from the original text can change interpretation quite drastically – 'Scientist' would indicate this is a discovery of scientific (healthcare most likely) significance; this is potentially a subjective influence we don't want, depending on the audience we are targeting the headline at. If scientific readership, then this is acceptable, however if it's a general headline, then the original headline would perhaps be more appropriate. It could be that the actual fossil was discovered by archaeologists, and scientists only subsequently discovered the significance of the cardiovascular system when the find was categorized and logged. Looking at financial news, for example detailing a merger or acquisition, the 'objects' of the headline (i.e. the companies in question) will likely bring with them a set of pre-existing knowledge, thereby influencing how the headline is interpreted, but also affecting what Dor calls "cognitive processing effort". Peramunetillek and Wong (2005) recognise this same phenomenon, suggesting that people will implicitly match their current situation with those situations in the past that are most similar to the current one, hence reading a news headline about an acquisition, which is similar to an historical entry, and that triggered a specific movement in the market (for example) could be a fair assumption and basis for impending activity.

A balance must then be struck in order to achieve the optimum processing effort by the reader, to ensure speedy interpretation and quality of content. This is further complicated by inclusion of contextual elements, specifically their positioning within the headline, and the cognitive processing capability of the end user. Clearly the latter of these three elements are not controllable in any practical sense, nor are the relationships between contextual elements, and how they increase cognitive processing requirements; what is controllable, and measurable, is the number of contextual words within a headline, which can be tested via experimentation.

2.1.6 Text summarization, and Natural Language Processing (NLP)

One area within technology that has attempted to address the challenge of automating news content is NLP, or Natural Language Processing, and the specific discipline of 'Text Summarization'. Thomson Reuters employs several different technologies which aim to summarize, abstract and generate text in order to update end products (such as [Thomson Reuters Eikon](#)). They do this by condensing the content of text & data, and then extract the most significant or pertinent facts or topics within them. The realm of NLP, AI and human-computer interaction is beyond the scope of this research paper, however it is related to the ultimate output of a best practice framework. Text summarization is based upon a set of rules, or learning algorithms, to produce an ongoing corpus that the AI can reference – machine learning allows the continual development of this corpus, to ensure relevancy and accuracy over time. A news headline best practice could easily be realized through the application of text summarization technology, learning over time what news headline content, structure and format are the most preferred, depending on the news story type and subject matter.

This research could form the basis for such an NLP system; we want to expose whether certain elements which make up the structure of a news headline are of significant value to editorial teams i.e. the identification of high priority best practice elements which significantly increase cognitive engagement.

2.2 Literature Review

Generally, the subject of news headline (and wider news story) editorial process best practice is largely unexplored – this is probably somewhat as a result of editorial teams and businesses wanting to maintain a competitive advantage, but it could also be due to the vast amount of influencers that affect news headline development. Most research in the public domain generally focuses on the problems with subjectivity, and how news headlines are adversely influenced.

Molek-Kozakowska (2013) proposes a 'pragma-linguistic' framework for the discovery of sensationalism in news headlines – although Molek-Kozakowska is not focusing on objectivity or best practice exclusively, the work does explore how headlines can be influenced, which is in a sense similar to my objective; to understand what elements have the most effect on news headline preference. Molek-Kozakowska (2014) also explores this concept further in her work on coercive metaphors in news headlines – she outlines various examples of where news headlines are either simplified, subject to imaging (metaphorical images such as 'it's an up-hill battle'), animalization, confrontation, (de)legitimization, emotionalization (emotional influence) and/or dramatization. If we look closer at the 'dramatization' example Molek-Kozakowska presents:

- Cosmetic face filler timebomb: Doctors call for crackdown over rising toll of women scarred by botched skin treatments.

Molek-Kozakowska states that "The words "toll" and "bomb" enable the readers to construct a mental image of an explosive device that, when placed under skin tissue by an incompetent practitioner, is likely to "go off" and kill or maim the patient at some point in time.". Far from being an objective interpretation of news content, the editorial team or individual behind the development of this headline, have intentionally sensationalized the story. Molek-Kozakowska concludes that "some figurative devices may have far-reaching consequences for framing the readers' interpretations of covered issues, and, in the long run, transforming their mental models of elements of socio-political reality"; review of past literature certainly corroborates this view, when considering Dor (2003) relevance optimizers, Zhang (2013) and his work into headline mediation, and Lindemann (1990) comments around subjectivity.

Chapter 3. THEORY

3.1 Hypotheses

As we have seen from the literature and existing research, news headline content (words), structure (ordering) and format (font) varies drastically depending on a wide variety of influencing factors; news subject matter, editorial process, political preference and general socio-economic elements all either directly or subjectively influence resulting news headline content. All of these elements ultimately affect interpretation speed, efficiency of processing, interpretation of the story itself and ability to act on information. Given the huge number of variables that may influence news headline composition, it is a challenge for editorial teams to standardize the approach to ensure consistency – it is my view that certain elements of news headlines can be extracted, measured and standardized via a best practice framework.

My initial top level hypotheses fall into three broad areas:

1. Headlines that contain descriptive words or phrases **AND** that contain contextual, relevant presupposed information **AND** that are designed as questions, statements or verb-phrases **OR** include names and concepts with high 'news value' achieve higher relevancy ratings than unstructured headlines.
2. Headlines containing, or adhering to, the Subject-verb-object triple structure **AND** that are concise, scannable and objective **AND** that contain 8 words or less **OR** that contain active verbs achieve stronger cognitive engagement than other unstructured headlines.
3. (Headlines that are formatted Verdana **AND** size 12 **AND** coloured black on a white background **AND** are hyperlinked are deemed easier to read (speed of reading), more legible, more attractive and achieve higher conversion rates than unformatted headlines.

Each of the above three hypotheses are designed to include several best practice elements, utilising an AND/OR structure, to demonstrate which elements have the most impact. Given the number of possible combinations of influencing variables, I have introduced a ranking system which identifies those 'high priority' influencers, or those variables which have the highest impact on headline structure improvement, based upon the research completed.

3.1.1 Headline Content (highest to lowest)

- a. **C1** Contain descriptive words, phrases
- b. **C2** Should contain contextual, relevant presupposed information
- c. **C3** Designed as questions, statements or verb-phrases
- d. **C4** Should include names and concepts with high 'news value' for the readers

3.1.2 Headline Structure (highest to lowest)

- e. **S1** Subject-verb-object triple
- f. **S2** Concise, scannable and objective
- g. **S3** Contain 8 words or less
- h. **S4** Contains active verbs

3.1.3 Headline Structure (highest to lowest)

- i. **F1** Formatted Verdana 12, coloured black on white
- j. **F2** Hyperlinks should connect the story to previously known facts and events

This provides us with a number of individual hypotheses to test; each hypotheses falls into one of the content, structure or format categories, enabling me to test the high priority elements against a variety of scenarios – this paper will look to test if the following statements are true -

1. Headlines that contain all high priority **content** BP (best-practice) elements achieve higher relevancy ratings than unstructured headlines, or those which contain low priority BP elements.
2. Headlines that contain all high priority **content** BP elements achieve higher relevancy ratings than unstructured headlines, or those which contain a single high priority BP element.
3. Headlines which contain a single high priority **content** BP element achieve higher relevancy ratings than unstructured headlines, or those which contain a single low priority BP element.
4. Headlines that contain all high priority **structural** BP elements achieve higher cognitive engagement than unstructured headlines, or those which contain low priority BP elements.
5. Headlines that contain all high priority **structural** BP elements achieve higher cognitive engagement than unstructured headlines, or those which contain a single high priority BP element.
6. Headlines which contain a single high priority **structural** BP element achieve higher cognitive engagement than unstructured headlines, or those which contain a single low priority BP element.
7. Headlines that contain all high priority **formatting** BP elements achieve higher conversion rates than unstructured headlines, or those which contain a single high priority BP element.

8. Headlines that contain all high priority BP elements are preferred to unstructured headlines, or those which contain less than the total recommended BP elements.

I would expect to find that if the high priority elements are included within a news headline, then additional inclusion of lower priority elements would have a minor positive impact upon overall news headline engagement, given the high priority items should provide the most cognitive cost-effectiveness. The 'best' headline is the one that adheres to all content, structure and formatting best practices, given application of just one category could result in a less efficient headline.

It is important that we are able to test the various best practice elements individually, as combining them into single tests would not enable us to evaluate how effective each element is when considering readers preferences.

For example, applying all Content best practice principles to our example news headline '**Sea creature fossil found with oldest-known cardiovascular system**' would result in something like the following -

Scientists discover 520 million year old shrimp-like sea creature fossil with oldest-known cardiovascular system.

This headline would clearly not satisfy Structure best practice items S2 'Concise, scannable and objective' nor S3 'Contain 8 words or less', but the inclusion of the Structure best practice would obscure overall preference, and not clearly indicate the 'strongest' elements.

Chapter 4. ANALYSIS AND DESIGN

4.1 Interviews Design

Interviews were designed to capture Thomson Reuters editorial staff preferences for developing news headlines – I targeted 5 individuals within the organization that had sufficient experience in the area of editorial editing, and who also worked within the editorial function of the business. Out of these 5 individuals, I was able to interview 3, with the remaining 2 not able to attend due to business commitments.

4.1.1 Description, basis & justification

The interviews with Thomson Reuters editorial staff is designed to be largely qualitative in nature, with only a few key questions common throughout each interview. The interviews were largely open ended, a deliberate design choice, given the editorial approach of each individual was specific to the individual, and not a part of any company wide best practice methodology.

There were several popular interview techniques to choose from, primarily falling into one of three categories; structured, semi-structured or unstructured. The structured approach would require direct, specific questions following a static order. This approach would have been difficult given the general approach to news headline composition is largely qualitative, and unstructured in nature – there is no industry standard to speak of. As a result, I believe an unstructured, free-flowing discussion with no specific questions set would be the most efficient method of gaining in depth responses, and provide the richest content. Winstanley (2014) neatly categories these three styles, with a brief description of approach – highlighted is the unstructured approach I took:

Structured	Semi-structured	Unstructured
Uses direct and specific questions only	Some key questions planned, with allowance for other issues to be raised	Free-flowing discussion; no fixed agenda

Specific order of questions	Indicative order of questions, but okay to depart from the order	No specific order for questions
The focus on how many people make the same points rather than individual views	Supplementary questions are offered to collate people's different viewpoints, but all are expected to answer the main questions	The interviewer is seeking depth of response and follows the interests of the interviewee
Must follow fixed schedule	Can leave out some questions as appropriate	Difficult to replicate as follows interests of interviewee and these will differ from person to person
Rather rigid style	Relaxed style	Conversational

Table 1 Interview Methods

4.1.2 Interview preparation

The primary objective of the unstructured interview is to gain an understanding of how Thomson Reuters editorial teams approached news headline development – I applied McNamara (2009) in Turner (2010) eight identified principles for interview preparation:

- **Choose a setting with little distraction;**
Interviews were conducted within an enclosed, sound proofed conference room on company premises.
- **Explain the purpose of the interview;**
An 'elevator pitch' was provided via email, and re-stated at the beginning of the interview for all participants.
- **Address terms of confidentiality;**
Participants were fully informed of the report confidentiality.
- **Explain the format of the interview;**
Participants were made aware of the unstructured approach to the interview.
- **Indicate how long the interview usually takes;**

In each case, 1 hour was assigned for the interview.

- **Tell them how to get in touch with you later if they want to;**

Participants were left with both an email and phone number to contact me on with any further insights, questions or queries.

- **Ask them if they have any questions before you both get started with the interview;**

Participants were asked whether they understood the objective of the interview, and whether they had any questions.

- **Don't count on your memory to recall their answers**

All answers were scribed via written notes.

4.1.3 Interview Participant selection

Participants were selected based upon their assumed knowledge in the area of news story editorial in accordance with a) their job title and b) and initial conversation to understand whether they were appropriate interview candidates. Creswell (2007) in Turner (2010) suggests the “importance of acquiring participants who will be willing to openly and honestly share information or “their story””. This is important given the content of news headlines is extremely sensitive, especially when considering Thomson Reuters staff who are a critical part of remaining competitive in the news marketplace.

4.2 Survey Design

The survey is designed to test the framework via the news headline output, based upon the literature research and interview output. Each survey question will present the reader with a news excerpt, along with several options for associated news headlines – the user in each case should choose which news headline is preferred, according to the particular

content, structure or format consideration. The following example demonstrates the structure of each question:

Q1 Edit Question Add Question Logic Move Copy Delete

1. (Reuters) - The U.S. Navy said on Tuesday it continued to examine options for a system to help Northrop Grumman Corp's new MQ-4 Triton high-altitude unmanned spy plane detect and avoid other aircraft, and expected to decide on a new approach in less than a year.

U.S. Navy considers options to ensure drones avoid other planes

U.S. Navy considers options to help new MQ-4 Triton spy drones avoid other planes

U.S. Navy mulls options to help drones avoid other planes

Figure 2 Survey Question

The initial news excerpt provides the background to the story, with the subsequent three choices being comprised of the original news headline, and two framework generated headlines utilizing my best practice principles.

Each option presented to the reader is randomly ordered, so as not to reveal the original news headline (in the public domain). The randomly ordered options also ensures the best practice elements I believe will be the preferred choice do not fall into a pattern.

I also include an open ended question, to allow participants to include more detailed information on their preference –

1. Considering digital news headlines, what motivates you to click through to the full story? Consider relevancy, format/appearance, time, device type and any other factor that may influence your behaviour.

This question is designed in part to capture post-survey recommendations on best practice element application, given future work recommendations would include a post-framework design second round of interviews (detailed further in section 7.4)

4.2.1 Description, basis & justification

The survey will be used to collect data around preferred best practice elements within news headlines, as determined by the developed framework. This allows us to test our hypotheses, and validate whether the best practice elements which we have identified (from the literature and interviews) has a significant impact upon end user preference.

For each question, I am including a small news excerpt from the full story – this is to help ‘standardise’ contextual, pre-supposed knowledge of a particular news story, given this particular element is not easily controllable. For example, should an individual taking the survey have deep knowledge around a particular news story, the headline they choose may well not contain significant factual information. In order to mitigate this effect, I am including a news excerpt and asking specifically which news headline is preferred in relation to that text. This will not entirely mitigate the pre-supposed issue, but it does significantly reduce the impact of the influence.

The survey is entirely anonymous, in accordance with Thomson Reuters privacy requirements – no personally identifiable information will be asked for, stored nor released. For the purpose of the survey, we will be using SurveyMonkey – this SaaS tool provides the most flexibility, and due the hosted nature is distributable to all identified participants.

4.2.2 Survey Participant selection

We focused the survey exercise on internal Thomson Reuters staff, in a variety of positions – the survey is entirely anonymous, hence the particular job function and area the individual resides is not known. As such, a mix of editorial and non-editorial respondents will be included, to represent general readership categories, as outlined within the interview results in section 5.1.1.

4.2.3 Use of priority ranking items in workflow

Each question & news headline will be designed according to the best practice framework output – given the number of best practice elements identified, I introduced a ranking sys-

tem to determine which elements had the most influence on the news headline. This ranking was derived from the literature research and Thomson Reuters interviews.

4.3 Framework

The best practice framework is the deliverable, responsible for outputting best practice news headlines. This artefact will be used to output the best practice news headlines used within the survey, with the survey itself being designed to specifically structure each headline according to the priority ranked best practice elements.

4.3.1 How the framework will be built

The framework will be built utilizing my existing knowledge of news headline development, the various influencing factors prevalent within the literature research, and finally via the interview output.

The framework will be designed in such a manner that technical specifications can be written based upon the functional output I will deliver – the rationale behind this approach is the hope that a second phase of this research, would be to build an application, which automates the appliance of best practice elements to news story text.

Chapter 5. METHODS AND REALIZATION

5.1 Approach

The research was approached in three phases; firstly, interviewing Thomson Reuters editorial staff in order to extract best practice elements from their working methods. This was compared with the best practice elements already identified from the literature, a comparison made and changes applied where appropriate. The next phase was the construction of the best practice framework. The third phase was testing the framework via a survey of news headlines – options which were determined by the framework best practice elements.

5.2 Interview Results

Below is a summary of the interview results, extracting key informational components which are most relevant to our set of hypotheses. The results have been combined into several sections, outlining the most significant influencing components. The data was collected via note taking during the interview discussion, as per the interview design outlined in the previous chapter.

5.2.1 Readership Focus

One of the artefacts delivered as a part of the interview(s), was a Venn Diagram depicting Thomson Reuters readership; this broke down end user consumption of information into three key areas – Professionals, Media Client and the Public:

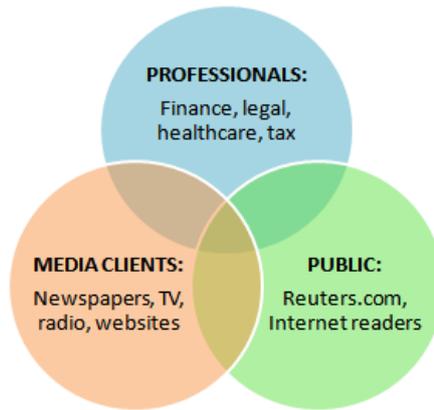


Figure 3 News Readership Categories

How an end user is categorized affects what kind of news content that individual is most interested in; for example, professionals are more interested (and likely to respond) to short, sharp factual information, data they can act upon quickly, whereas the Public category depicts a more casual user, perhaps not as time sensitive. The type of end user clearly has an impact on the type of news delivery and structure, however this particular element was not in scope of this paper's research – this would form perhaps a second phase of the report, to understand how the type of end user may additionally influence news headline development.

5.2.2 Thomson Reuters 'Do's and Don'ts'

The TR elevator pitch is "*To produce clear, eye-catching headlines that draw readers to our stories*" – in order to remain true to this statement, one interviewee revealed a "Headlines Do and Don't" list, which is informally utilised throughout the editorial writing teams work:

GOOD HEADLINES	WEAK HEADLINES
<ul style="list-style-type: none"> • Short, sharp, focused • Consistent with tone, facts of story • Short words, active verbs • Present tense for action • A milestone, ranking, touch of gold • Changes as story moves • Maximum of 64 characters including tags 	<ul style="list-style-type: none"> • Too many ideas, cluttered • Odd abbreviations, obscure acronyms • Jargon, cliches, journalese • Long titles for people • Unclear sourcing • Location unknown • Too long

Figure 4 Interview Headline Recommendations

Many of these elements are similar to our identified best practice elements that have been largely derived from my prior research of existing literature. Certainly the “short, sharp, focused” and “short words, active verbs” items speak to the subject-verb-object structure, along with the concise, scannable and objective elements that Morkes & Nielsen (1998) outline.

The participants also went into a little more detail around subsequent news story content, once the initial news headline had been delivered. This was described as a news story “half life”, or the amount of time the story ‘lives’ before interest in the subject matter begins to dissipate. This can have an impact upon subsequent stories and news headlines released in relation to the original breaking story, in many cases changing the type of content. For example, consider the below breaking news headline:

- Australia's David Jones (DJS.AX) jumps 23% on South Africa's Woolworths takeover bid, valued at \$2.15 Billion

A subsequent headline and story related to the original story may read something like the following:

- DJS AX up 23% on takeover bid, valued at \$2.15 Billion

Key information such as named entities, 'active' verbs (jumps) and other components have been removed given information may be current enough in the readers mind not to require repeating. This would not always be the case, but as headlines age, the 'value' of that headline and story states to decrease. One participant provided the following headline categories and descriptions, in relation to the idea of news story life:

- **Urgent headlines** - should be clear, free of jargon and informative. These are often the primary released headline when the story is breaking, which needs to provide the most information, or 'bang for your buck' in the least amount of time.
- **Update 1+ headlines** – should show why the story is important. Q1, economic abbreviations, too many numbers etc are unwelcome. The news story example above speaks to this kind of story development, where subsequent updates are released over the lifetime of a news headline.
- **Wrap-up** headlines – should go beyond the basic facts and offer something more, making readers want to read on. This is often in relation to a 'long-life' story, for example, a current global conflict, a pandemic or some other wide reaching, 'big' story. This is where subjective influences can manifest themselves, as the story develops.

There was also some focus around "to pun or not to pun", in the sense does introducing a pun to a news headline either increase or decrease reader engagement? A few examples cited –

- **Foot off the gas? Investors fret over Hyundai's long-term growth**
SEOUL - Hyundai Motor Co drove up sales in France by 40 percent last month and had a record September in the United States, but it has since had nearly \$7

billion wiped off its market value as investors query the South Korean auto-maker's go-slow on expanding capacity

- **Indonesia digs hole for itself with new mining laws**

JAKARTA - For decades explorers have enjoyed a place on the cutting edge of Indonesia's mining boom as they scoured the earth for fresh mineral deposits. But now they complain that new rules aimed at boosting state revenue from natural resources have slashed investment in mineral prospecting and could threaten the entire \$93 billion sector.

- **Investors look to water for a steady flow of returns**

ZURICH/LONDON - Investors disillusioned with rock-bottom interest rates and volatile equity markets are tapping into the global water market, as they bet the growing need for fresh water supplies could transform the sector into a money-spinner to rival oil.

The general consensus around “punning” was largely “don’t”. Although it can bring some brevity to light-hearted stories, it otherwise reduces the effectiveness of the overall news story, and implies a non-serious tone. This is not a particularly big problem for the individual news headline, but more the influencing, contextual affect it has upon the wider news company – the publishing news organizations professional reputation could potentially be tarnished.

5.2.3 Editorial Authorship

An interesting area, and potential barrier to a framework being widely utilized, is the concept of editorial right to authorship. All interviewees expressed concern around a best practice framework that may reduce the journalist or editorial teams ability to author a headline for a news story – a kind of infringement on ‘creative license’. This in itself presents an issue for objective news headline development. The idea behind our best practice framework is to facilitate the editorial process, augmented existing capability in order

to assist in headline generation. News publishing technology has now advanced sufficiently to allow journalists in the field the ability to publish stories directly, without a central editorial team's intervention. Here, I would argue, a best practice framework automation could help in identifying potentially subjective news headlines, filtering content and delivering objective stories to the end user.

5.3 Interview output and Utilisation

The interview output will be utilized via a comparison with the best practice elements identified through the literature research. We hope to find, in many cases, similarities between what has been identified in prior research, and what is being practically applied by professionals in the field today. Observing a correlation between these two aspects will strengthen our overall extracted best practice elements.

5.4 Survey

The survey will be implemented using an online SaaS survey tool, SurveyMonkey. The results will then be tested against the initial hypotheses, to understand whether the best practice elements I have chosen do in fact make a significant impact on readership engagement.

5.4.1 Survey Implementation

Using SurveyMonkey, I will be designing multiple choice questions, designed to test the readers preference for headline structure – the below demonstrates 5 options for the reader to consider, including the original news headline.

Q6 Edit Question ▼ Add Question Logic Move Copy Delete

6. (Reuters) - The hot question of whether dinosaurs were warm-blooded like birds and mammals or cold blooded like reptiles, fish and amphibians finally has a good answer.

Dinosaurs; warm or cold blooded?

Dinosaurs, wiped out by an asteroid 65 million years ago, were both warm and cold blooded.

Biologist John Grady, in a study of 21 dinosaurs including Tyrannosaurus Rex, reveals new metabolic data

Warm blooded or cold? Scientists uncover new data revealing they were somewhere in between

Warm blooded or cold? Dinosaurs were somewhere in between

Figure 5 Survey Question

5.4.2 Survey results analysis method

Analysis of the survey results will be quantitative in nature. Statistically, I will look at the overall preference rankings, which should determine which best practice elements are the most preferred, when considering both a mix and individual components.

5.5 Framework Design

Best practice framework design according to identified priority/secondary elements.

5.5.1 The framework design

The below framework design workflow exhibits the identified best practice priority elements along with the secondary elements e.g. Content elements C1 and C2 are priority ‘must have’ elements for news headlines, whereas I recommend just one of Content C3 or C4 elements in addition to the priority items, given they are lower priority elements.

Hence, an editorial staff member would begin with the raw news story data, and run through the workflow, applying the particular best practice at each stage. Content C1 and C2, Structure S1 and S2 and Format F1 and F2 must be applied in each case, with Content C3 and C4 and Structure S3 and S4 being optional elements for inclusion.

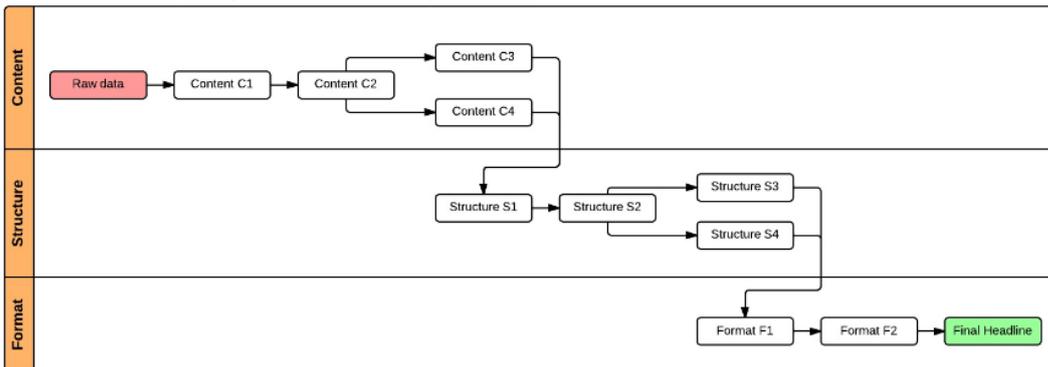


Figure 6 Framework Design

Remembering our three top level hypotheses, I indicated that priority best practice elements e.g. C1 and C2 must be included, along with one low priority element i.e. C3 or C4; the addition of a single low priority best practice element to an already filtered news headline (that has had high priority items applied) should positively affect the headline, but not as significantly as the initial high priority element application.

5.5.1.1 Framework - Content

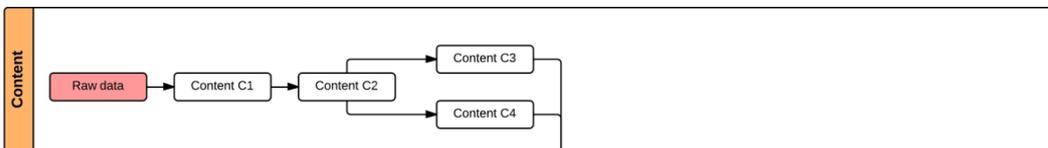


Figure 7 Framework Content Design

1. **Headline Content (highest to lowest)**

- a. **C1** Contain descriptive words, phrases **AND**
- b. **C2** Should contain contextual, relevant presupposed information **AND**
- c. **C3** Designed as questions, statements or verb-phrases **OR**
- d. **C4** Should include names and concepts with high 'news value' for the readers

Starting with the raw data i.e. the news story content, the editorial team or individual can approach headline generation in one of two ways. Firstly, produce a 'working' headline and then run that headline through the content best practice elements, or secondly, gen-

erating a news headline from scratch utilizing each best practice item at each step of the process workflow. For the first approach, if we consider our example news story ‘Sea creature fossil found with oldest-known cardiovascular system’ and rank it against the best practice content, we can check whether the headline already contains those elements – if not, the headline can be modified at each step to include those elements. For the second approach, editorial teams take the content of the news story, and build a headline according to each best practice element step by step.

5.5.1.2 Framework - Structure

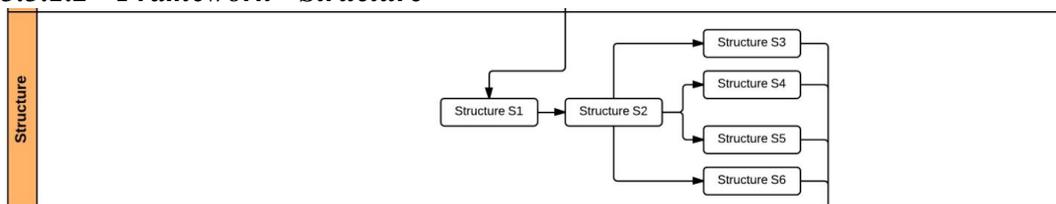


Figure 8 Framework Structure Design

1. Headline Structure (highest to lowest)

- a. **S1** Subject-verb-object triple **AND**
- b. **S2** Concise, scannable and objective **AND**
- c. **S3** Contain 8 words or less **OR**
- d. **S4** Contains active verbs

Following the content best-practice application, the structural best practice elements are applied in the exact same manner, utilizing the work-flow in order to augment and existing ‘working’ headline, or to continue generation of a new headline through application of each best practice element step by step.

5.5.1.3 Framework - Format



Figure 9 Framework Format Design

2. Headline Structure (highest to lowest)

- a. **F1** Formatted Verdana 12, coloured black on white **AND**

- b. **F2** Hyperlinks should connect the story to previously known facts and events

Finally, application of the format best practice (following content and structure) – although I was not able to test the hyperlinks best practice element F2, research and interview data points to this functionality being highly preferred over ‘dumb’ text i.e. non-hyperlinked headlines.

Chapter 6. RESULTS AND EVALUATION

6.1 Survey Results

We received a total of 36 responses between the 28th of June to the 10th of July – the peak response period on the 30th June is mainly due to the weekend, given the survey respondents were Thomson Reuters employees, and would not have seen the survey link via the internal intranet site until Monday AM (30th June 2014).

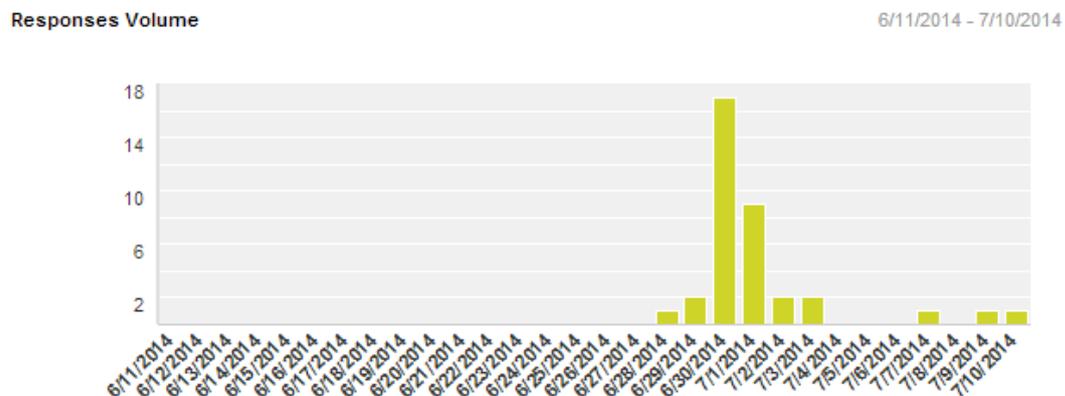


Figure 10 Survey Responses

6.2 News headline criteria, a refresh

First let us remind ourselves of the criteria applied to each news headline – in all three categories (content, structure & format) we ranked best practice criteria; a summary of these criteria is shown below. Survey participants then chose a preference from the list of headlines, which contained various combinations of the below elements -

1. *Headline Content (highest to lowest)*

- a. **C1** Contain descriptive words, phrases
- b. **C2** Should contain contextual, relevant presupposed information
- c. **C3** Designed as questions, statements or verb-phrases
- d. **C4** Should include names and concepts with high 'news value' for the readers

2. **Headline Structure (highest to lowest)**

- a. **S1** Subject-verb-object triple
- b. **S2** Concise, scannable and objective
- c. **S3** Contain 8 words or less
- d. **S4** Contains active verbs

3. **Headline Structure (highest to lowest)**

- a. **F1** Formatted Verdana 12, coloured black on white
- b. **F2** Hyperlinks should connect the story to previously known facts and events

6.3 Results overview, by hypotheses

- 1. Headlines that contain all high priority **content** BP (best-practice) elements achieve higher relevancy ratings than unstructured headlines, or those which contain low priority BP elements.

Content Test 1					
	Original	c1c2	c3c4		
Scientific	19.4%	22.2%	58.3%		
Scientific	22.9%	25.7%	51.4%		
Financial	45.5%	36.4%	18.2%		
Financial	18.2%	50.0%	31.8%		

Table 2 Content Test 1

Note - For each test, I included two separate headline tests in each news area i.e. 2 within Scientific and 2 within Financial – this is to reduce the risk of a headline being known by the survey participant, and to also increase the sample size of results.

For the scientific news headlines, user preference was for best-practice elements C3 and C4, whereas my initial hypotheses had stated that C1C2 were the most effective when considering headline content. The story changes for the financial news headlines, where the preference was for a mix of the original headline and the C1C2 mix, as we had origi-

nally suggested. 1 out of 4 scenarios showed a preference for the C1C2 mix. **Hypothesis false.**

- Headlines that contain all high priority **content** BP elements achieve higher relevancy ratings than unstructured headlines, or those which contain a single high priority BP element.

Content Test 2					
	Original	c1c2	c1	c2	
Scientific	8.3%	44.4%	11.1%	36.1%	
Scientific	25.0%	52.8%	8.3%	13.9%	
Financial	13.6%	72.7%	4.5%	9.1%	
Financial	18.2%	63.6%	0.0%	18.2%	

Table 3 Content Test 2

For Content Test 2, all respondents reported a preference for the hypothesized C1C2 mix of best practice elements. This in combination with hypotheses 1 indicates a preference for the high priority best practice elements. 4 out of 4 scenarios showed a preference for the C1C2 mix. **Hypothesis true.**

- Headlines which contain a single high priority **content** BP element achieve higher relevancy ratings than unstructured headlines, or those which contain a single low priority BP element.

Content Test 3					
	Original	c1	c2	c3	c4
Scientific	8.3%	27.8%	44.4%	0.0%	19.4%
Scientific	22.9%	22.9%	0.0%	54.3%	0.0%
Financial	50.0%	0.0%	4.5%	9.1%	36.4%
Financial	18.2%	4.5%	45.5%	13.6%	18.2%

Table 4 Content Test 3

For Content Test 3, there was no clear preference for the individual best practice elements – this would suggest that each content best practice item, in isolation, was not as effective as a combination thereof. 2 out of 4 test scenarios showed a preference for either C1 or C2 elements. **Hypothesis indeterminate.**

4. Headlines that contain all high priority **structural** BP elements achieve higher cognitive engagement than unstructured headlines, or those which contain low priority BP elements.

		Structure Test 1			
		Original	s1s2	s3s4	
Scientific		18.5%	33.3%	48.1%	
Scientific		22.2%	44.4%	33.3%	
Financial		40.0%	40.0%	20.0%	
Financial		55.0%	35.0%	10.0%	

Table 5 Structure Test 1

For the first Structure test, results suggest that the high priority S1S2 mix were slightly preferred over the original and S3S4 mix; no clear patterns emerged concerning financial versus scientific news headlines. 2 out of 4 test scenarios showed a preference for the S1S2 mix. **Hypothesis indeterminate.**

5. Headlines that contain all high priority **structural** BP elements achieve higher cognitive engagement than unstructured headlines, or those which contain a single high priority BP element.

		Structure Test 2			
		Original	s1s2	s1	s2
Scientific		33.3%	40.7%	7.4%	18.5%
Scientific		11.1%	33.3%	44.4%	11.1%
Financial		26.3%	15.8%	31.6%	26.3%
Financial		68.4%	10.5%	15.8%	5.3%

Table 6 Structure Test 2

The second Structure test did not show any strong preference for the S1S2 mix; in each of the 4 scientific and financial news headlines, a different preference was selected each time. 1 out of 4 test scenarios showed a preference for the S1S2 mix. **Hypothesis false.**

6. Headlines which contain a single high priority **structural** BP element achieve higher cognitive engagement than unstructured headlines, or those which contain a single low priority BP element.

		Structure Test 3				
		Original	s1	s2	s3	s4

Scientific	22.2%	33.3%	11.1%	11.1%	22.2%
Scientific	29.6%	11.1%	22.2%	7.4%	29.6%
Financial	31.6%	5.3%	47.4%	15.8%	0.0%
Financial	15.8%	36.8%	21.1%	21.1%	5.3%

Table 7 Structure Test 3

For Structure Test 3, a preference can be observed for S1 and S2 elements, with the financial news headlines returning the most positive result. 3 out of 4 test scenarios showed a preference for either S1 or S2 elements. **Hypothesis true.**

- Headlines that contain all high priority **formatting** BP elements achieve higher conversion rates than unstructured headlines, or those which contain a single high priority BP element.

		Format Test 1			
		f1f2	f1	f2	
Scientific	73.1%	15.4%	11.5%		
Financial	61.9%	23.8%	14.3%		

Table 8 Format Test 1

In all cases for the Format Test, the preference was clearly indicated as the F1F2 mix, as our initial hypotheses had outlined. 2 out of 2 test scenarios showed a preference for the F1F2 mix. **Hypothesis true.**

6.4 Overall analysis, financial and scientific

When reviewing the results combined, a pattern starts to emerge around best practice element preferences – below is a summary count of each element/element mix, by number of preferences, not accounting for news area (scientific & financial combined).

Note: for the purpose of this analysis, we consider the original headline in the category of best-practice “mix” e.g. C1C2 or C3C4 or Original. For example, one of the tests original headline reads – “U.S. Navy mulls options to help drones avoid other planes”, with two subsequent options which present a mix of best practice elements (C1C2 or C3C4). The reason for this is due to the original headline in some cases contains elements of best practice mix already, which is a preference we want to capture.

All			
Combination	Count	Combination	Count
C1C2	5	S1	4
Original	2	S1S2	2.5
C3C4	2	Original	2.5
C2	2	S3S4	1
C3	1	S2	1
C1	0	S3	0.5
C4	0	S4	0.5

Table 9 Overall Analysis Financial & Scientific

The results above translate into the following '1st' and '2nd' place rankings below, which is largely in-line with our initial broader hypotheses around content, structure and format.

Results (actual)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	Original
Content Preferred Isolated Element	C2	C3
Structure Preferred Mix	S1S2	Original
Structure Preferred Isolated Element	S1	S2

Table 10 Overall Analysis Financial & Scientific, Actual

Results (hypotheses)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	C3C4
Content Preferred Isolated Element	C1/C2	C1/C2
Structure Preferred Mix	S1S2	S3S4
Structure Preferred Isolated Element	S1/S2	S1/S2

Table 11 Overall Analysis Financial & Scientific, Hypotheses

The preferred overall mix for content was the C1C2 mix, with the original news headline being the secondary preference. The same is true for the structural S1S2 mix, again with the original headline being the secondary preference. When looking at individual elements, for content, C2 was preferred over C3, where our hypotheses would have expected the C1 element to be in one of the top spots. For structure, our hypotheses holds, with both S1 and S2 elements holding the top two spots.

The preference for the original news headline in 4.5 cases (the 0.5 is where the ranking was jointly shared by another preference) suggests that the original contained a better, or more appropriate, mix of best practice elements than the revised headline. The headlines in question are as follows –

- Australia's David Jones jumps 23 percent on South Africa's Woolworths takeover bid (**financial**)
- Sports Direct share price sinks after founder cuts stake (**financial**)
- HTC shares fall after bigger-than-expected first quarter loss (**financial**)
- Son of senior Shi'ite cleric sentenced to death in Saudi Arabia (**financial**)
- Shaking Shakespeare: Richard III was no hunchback after all (**scientific**)

Of the 4.5 original headline preferences, 4 were from the financial news stories, with just 0.5 from the scientific news stories (shared with structure best practice element S4). In order to understand why the original headlines were chosen over our best practice developed headlines, I review below which best practice elements already existing within those chosen headlines -

Headline	C1	C2	C3	C4	S1	S2	S3	S4	BP
1	Y	Y	Y	Y	Y	Y	N	Y	7
2	Y	N	Y	Y	Y	N	N	N	4
3	Y	Y	Y	Y	Y	N	Y	N	6
4	Y	N	Y	Y	Y	N	N	Y	5
5	N	N	Y	Y	N	N	N	N	2

Table 12 Original Headline Preference Analysis

The 4 financial headlines (1 – 4) exhibit a high level of best practice elements already in place (4 – 7 elements), which could explain the preference for the original version. The scientific headline only exhibits 2 elements, which could be the reason behind the original headline sharing 1st place preference with the S4 best practice element. Within the financial headlines, there is a consistent pattern in that elements C1, C3, C4 and S1 are always present – the survey results somewhat match this pattern, in that elements C1 and S1 are ‘top’ scoring elements, however elements C3 and C4 score low in most cases.

6.5 Overall analysis, financial versus scientific

Next, we will perform the same analysis as above, but filtering by news area i.e. scientific and financial. When looking at the results through these individual lenses, the picture changes somewhat, revealing a preference for best practice within the scientific area, but not so strongly when considering financial news headlines:

Note: for the purpose of this analysis, we consider the original headline in the category of best-practice “mix” e.g. C1C2 or C3C4 or Original. For example, one of the tests original headline reads – “U.S. Navy mulls options to help drones avoid other planes”, with two subsequent options which present a mix of best practice elements (C1C2 or C3C4). The reason for this is due to the original headline in some cases contains elements of best practice mix already, which is a preference we want to capture.

Scientific			
Combination	Count	Combination	Count
C1C2	2	S1	2
C3C4	2	S1S2	2
C2	1	S3S4	1
C3	1	Original	0.5
Original	0	S4	0.5
C1	0	S2	0
C4	0	S3	0

Table 13 Overall Analysis Financial Versus Scientific

The results above translate into the following ‘1st’ and ‘2nd’ place rankings below, which aside from Content item C3 and Structure item S4, is in-line with our initial broad hypotheses around content, structure and format.

Results (actual)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	C3C4
Content Preferred Isolated Element	C2	C3
Structure Preferred Mix	S1S2	S3S4
Structure Preferred Isolated Element	S1	S4

Table 14 Overall Analysis Financial & Scientific, Actual

Results (hypotheses)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	C3C4
Content Preferred Isolated Element	C1/C2	C1/C2
Structure Preferred Mix	S1S2	S3S4
Structure Preferred Isolated Element	S1/S2	S1/S2

Table 15 Overall Analysis Financial & Scientific, Hypotheses

The preferred scientific content mix was C1C2 compared to the C3C4 mix, with element C2 being the preferred isolated element. We would have expected element C1 to be the next preferred element, as opposed to C3. For structure, the preferred scientific mix was S1S2 compared to the S3S4 mix, with element S1 being the preferred isolated element. Element S4 came through as the preferred secondary isolated element.

Next, we look at the financial results breakdown:

Financial			
Combination	Count	Combination	Count
C1C2	3	Original	2.5
Original	2	S1	2
C2	1	S2	1
C3C4	0	S1S2	0.5
C4	0	S3S4	0
C3	0	S4	0
C1	0	S3	0

Table 16 Overall Analysis Financial

The results above translate into the following ‘1st’ and ‘2nd’ place rankings below; although the content mix is as we expected, with C1C2 holding the top spot, for structure, the original news headline was preferred over our best practice mix, even though the S1S2 mix came as the secondary preference.

Results (actual)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	Original
Content Preferred Isolated Element	C2	NA
Structure Preferred Mix	Original	S1S2
Structure Preferred Isolated Element	S1	S2

Table 17 Overall Analysis Financial, Actual

Results (hypotheses)

Combination	Primary	Secondary
Content Preferred Mix	C1C2	C3C4
Content Preferred Isolated Element	C1/C2	C1/C2
Structure Preferred Mix	S1S2	S3S4
Structure Preferred Isolated Element	S1/S2	S1/S2

Table 18 Overall Analysis Financial, Hypotheses

The preference for the original news headline within the financial area suggests that best practice elements are less effective when considered against scientific headlines. This does make sense considering the audience taking the survey – scientific headlines are by

their nature generally less time sensitive than financial headlines, where end users may be looking to utilize the information immediately. Scientific headlines are generally more narrative driven, with financial headlines being more informational (statistics, key entities, quantitative intelligence). An alternative explanation could be due to the survey respondents having been previously exposed to the Reuters editorial principles – although the Reuters approach is largely qualitative (and not part of any existing formal framework), participants are part of the Reuters culture on a day to day basis, and unconsciously may have influenced their headline preference.

6.6 Overall analysis, format

For the format tests, the results were significant and conclusive – we had stated that the F1F2 best practice mix should be the preferred option when given a number of formatting options, and in all case the F1F2 was clearly the preferred option. We had expected best practice element F1 to come in second place, which again was corroborated by the survey results.

All	
Combination	Count
F1F2	2
F1	2
F2	0

Table 19 Overall Analysis Format

6.7 Open ended question responses

As a part of the survey, as mentioned in the interview design (Survey design 4.2), I included an open ended question to capture further detail on news headline preference. Each response is listed below, along with identified best practice elements called out by the participant -

1. The headline has to be **relevant** to my interest, but it has to be **informative**, otherwise I treat it as spam.
 - 1.1. This speaks to best practice elements C1 and C2. The ‘relevance’ piece is something that is prevalent in both the literature review and interviews, and often coincides with the headline needing to be ‘informative’.

2. **Activity**, focus and insight.
 - 2.1. This speaks to best practice element S4 – a headline which is ‘active’ in the sense of active verb usage. The literature backs this view up, recommending active over passive verb usage.

3. The headline is an implicit promise to the reader: I will click on it if I think the story is going to be **interesting, lively** and **important**, or hopefully a combination of all three. A caveat on the survey - in some of your examples the 'most accurate' headline would not work in practice because it is just too long. And in some cases there are better alternatives than any of those listed.
 - 3.1. This speaks to best practice elements C1, C2 and C4. The participant here is specifically looking for ‘active’ headlines, those which pique an interest, relevant to the user. This is perhaps different from a more financial focused headline, where individuals are actively searching for information that may be commercially useful, as opposed to something which interests them.

4. **Not too long**, but also **precise**: I think I'm more likely to click on **something that has a number**, for example, rather than something too general. Also, I deeply **dislike emotional appeals**...
 - 4.1. This speaks to best practice elements S2, S3, C1 and C4. This participant makes a point of calling out emotional appeals, or in other words those which are objective, and are not influenced subjectively.

5. I don't need all the facts in the headline, otherwise there is nothing tempting me to click through. there should be some **credible hint** as to the nature of the story underlying the headline
 - 5.1. This speaks to best practice elements C1 and C2 – including contextual, pre-supposed information is likely to increase click through rates, as this participant outlines.

6. **Human interest**, feeling that there's a news story that I will remember and take something from either - either to **talk about** or to **enhance my knowledge of the world**.
 - 6.1. This speaks to best practice elements C1, C2 and C4. This participant states that a 'human interest' in a headline is of importance, something which is relevant to their interests.

7. **Brevity, large size, captures relevant topic or content of story (subject, verb/predicate, object)**, maybe a **brief abstract** (one sentence) when space permits, definitely not misleading.
 - 7.1. This speaks to best practice elements S1, S2, S3, F1, C1 and C2. This participant clearly has a prior knowledge of editorial process, given they specifically call out the subject-verb-object structure.

8. Interest in the topic
 - 8.1. This speaks to best practice elements C2

9. Highlighted (sic) information, **emotive commentary** (not just facts) obviously **missing information** (I have to click through to find out what I really want)
 - 9.1. This speaks to best practice elements C1, C2 and C3 – here the participant calls out the need for a hook, something to tempt the user to click through to the full story. This would perhaps be more relevant for scientific content, given many financial users rely on headlines alone to convey time-sensitive data.

10. **Short and snappy, easy to read**.
 - 10.1. This speaks to best practice elements S2, S3 and F1. The 'easy to read' piece is presumably speaking to a specific format, as opposed to the type of content (this would obviously need validation).

11. **Promise of insight** (i.e. if the headline gives me all the facts then I might not gain anything from clicking through). If I'm on a mobile device then I'll only click through if the **text is formatted and big enough to read**, otherwise I'll remember to read later when I'm on a bigger screen.

11.1. This speaks to best practice elements C2 and F1. The 'promise of insight' is a recurring theme in the comments, essentially pointing to a reason to click through the headline to the full story.

12. The headline needs a **'hook'** ... something that makes the story sound compelling.

12.1. This speaks to best practice elements C2. Again, we see an emphasis on a 'hook' (specifically called out here in name); participants generally seem to want the promise of more information after viewing the initial headline.

13. **Shorter headlines in a clear sentence**, rather than a mess of nouns; **slight ambiguity or wordplay** but only if supported by the story itself; avoid uncanny valley. If a standfirst is included with the headline, that is often what motivates me to navigate to the full story

13.1. This speaks to best practice elements S2, S3 and C2. This participant calls out the need for 'slight ambiguity', which is interesting given most of the literature recommends objective, clear headlines that do not contain any uncertainty around story content.

6.8 Hypotheses testing summary

I have analyzed the results compared with our initial hypotheses in three different ways – firstly assessing the results of each question against the relevant hypotheses, secondly looking at the overall preference by ranking (number of times a best practice mix or element is the overall preferred choice), and thirdly looking at the overall preference by ranking, by news area (financial versus scientific).

- In the first assessment, we find 3 out of 7 hypotheses tests are true, with 2 indeterminate and 2 false. The indeterminate tests are where results fall into a 50/50 scenario.
- In the second assessment, we review overall results ranking, which supports our hypotheses for best practice mix and isolated elements.
- In the third assessment, we review financial versus scientific preferences; the scientific analysis supports our hypotheses, however the financial headlines demonstrate a heavier preference for the original news headline

In both the second and third assessment, in all scenarios except for the financial structure best practice mix, the preference was for a news headline which had been edited according to our framework recommendations, as a result of our hypotheses.

The open ended comments also largely support our results; in the majority of cases participants support the priority elements inclusion with the headlines (C1C2, S1S2, F1F2), and also call out preference for the secondary elements. There were some elements mentioned which were not added as a part of our best practice framework, specifically more emotive, subjective elements, and some calling for a certain amount of ambiguity in the headline, to encourage click through. Given the objective of this research is to develop a framework which promotes 'consistent and objective' headlines, I consider these items to be out of scope.

Chapter 7. CONCLUSIONS

7.1 Lessons Learned

Additional best practice elements need testing, including both isolated and mixed sets. What is clear from the survey results, including the open ended questions, is that elements I have considered secondary to the priority items are perhaps more important depending on the context of the story. The first set of results provide us with an indication of preference, however further tests are required which encompass a wider set of news headline subject areas.

7.2 Academic Application and Limitations

There is a clear synergy between his research and authors of similar work, such as Molek-Kozakowska (2014), Dor (2003), Zhang (2013) and Lindemann (1990) – further research is required in order to more accurately define best practice elements for news headline construction, especially around the area of subjective influence. To this end, the academic application of this work is limited in its current state, however if augmented and further developed with the above authors work in mind, a comprehensive best practice framework could be developed.

7.3 Business Application and Limitations

The business application of this research is perhaps more beneficial than the academic prospects – given a large component of the final best practice elements was influenced by Thomson Reuters employee interviews, the framework is essentially a formalized version of what many editorial teams members were already applying. In that sense, the framework can be used to provide training and development to new starters, or the formation of editorial teams in news functions. The framework is of course still limited by the editorial function itself.

7.4 Recommendations for Future Research & Work

A new survey focusing on all best practice elements individually, along with a broader sample from multiple sources, in addition to Thomson Reuters personnel – given the respondents were from Thomson Reuters, commenting on Thomson Reuters news headlines, it is possible this may have influenced results. I would also want to see the framework utilized in commercial news headline development, with metrics being collected to measure the effectiveness of the framework.

The generated framework was the result of both research and interviews, with a subsequent survey testing my hypotheses – I would have liked to have carried out a second interview series in order to gain a more in depth feedback on the results, to understand why certain headlines were chosen, especially in cases where original headlines were the preferred choice. It is likely that several rounds of fine tuning of the framework is needed, in order to arrive at a more accurate and efficient model.

I also intend to develop a Proof of Concept application (Java based) with the objective of digitizing the best practice framework I developed – in practice, this would involve entering news story text OR a working news headline, working through the various steps in the framework, with the application making recommendations as to modification or structure. This would involve a deep level of Natural Language Processing capability, to build out an algorithm that could learn which best-practice elements were the most effective, via some kind of active feedback system.

REFERENCES CITED

1. Baggini, J (2002) *Making sense: Philosophy behind the headlines*, London: Oxford University Press
2. Bergstrom, G (2014) *Writing Hard headlines and Intros* [Online] Available: <http://marketing.about.com/od/publicrelation1/a/Writing-Hard-Headlines-And-Intros.htm> [20th February 2014]
3. Bernard, M, Lida, B, (2002) *A comparison of popular online fonts: Which size and Type is best?*, Usability News, 4(1), pp.6-7
[http://www.tbook.it/download/Usability_News%20.pdf]
4. Boczkowski, P (2005) *Digitizing the news: Innovation in online newspapers*, Cambridge, MA: The MIT Press
5. Caumont, A (2013) *12 trends shaping digital news* [Online] Available: <http://www.pewresearch.org/fact-tank/2013/10/16/12-trends-shaping-digital-news/> [20th February 2014]
6. Chua, R (2014) *(Re)Structuring Journalism* [Online] Available: <http://structureofnews.wordpress.com/2014/02/10/structured-ezra/> [20th February 2014]
7. Develotte, C, Rechniewski, E (2103) *Discourse analysis of newspaper headlines: a methodological framework for research into national representations* [Online] Available: <http://wjfms.ncl.ac.uk/titles.htm> [20th February 2014]
8. Dor, D (2003) *On newspaper headlines as relevance optimizers*, Journal of Pragmatics, 35(5), pp.695-721
9. Edelstyn, S (2013) *5 tips to help improve your headline click-through rate* [Online] Available : <http://www.theguardian.com/media-network-outbrain-partner-zone/5-tips-headline-click-through-rate> [20th February 2014]
10. Halbrooks, G (2014) *Writing Headlines for the Web* [Online] Available: <http://media.about.com/od/mediawritingtips/qt/Writing-Headlines-For-The-Web.htm> [20th February 2014]
11. Howard Bear, J (2014) *How to Write Effective Headlines* [Online] Available: <http://desktoppub.about.com/cs/techwriting/a/headlines.htm> [20th February 2014]

12. Judy, G (2004) *Writing for the web versus writing for print: are they really so different?*, *Technical Communication*, 51(2), p275-285
13. Juicy Studio (2009) *Readability Test* [Online] Available from:
<http://juicystudio.com/services/readability.php> [20th February 2014]
14. Khurana, A (2014) *A Better headline than this would have made you buy something from me!* [Online] Available: <http://ecommerce.about.com/od/Ecommerce-Sales-Copy/fl/A-Better-Headline-Than-This-Would-Have-Made-You-Buy-Something-From-Me.htm> [20th February 2014]
15. Kilian, C (2009) *Writing for the web*, 4th edition, Self-counsel Press
16. Matheson, D (2004) *Weblogs and the epistemology of the News: Some trends in Online Journalism*, *New media & Society*, 6(4), pp.443-468
17. Meifang Zhang (2013) *Stance and mediation in transediting news headlines as paratexts*, *Perspectives: Studies in Translatology*, 21:3, 396-411
18. Morkes, J, Nielsen, J (1997) *Concise, SCANNABLE, and objective: How to Write for the Web* [online] available: <http://www.nngroup.com/articles/concise-scannable-and-objective-how-to-write-for-the-web/> [20th February 2014]
19. Morkes, J, Nielsen, J (1998) *Applying writing guidelines to web pages*, *CHI 98 Conference Summary on Human Factors in Computing Systems*, pp.321-322
20. Molek-Kozakowska, Katarzyna (2014) *Coercive Metaphors in News Headlines: A cognitive-pragmatic approach*, *Brno Studies in English*, 40(1)
21. Molek-Kozakowska, Katarzyna (2013) *Towards a pragma-linguistic framework for the study of sensationalism in news headlines*, *Discourse & Communication* 7(2), pp. 173-197
22. Nielsen, J (2009) *Writing for the Web: Research on how users read on the Web and how authors should write their Web pages* [Online] Available from:
<http://www.useit.com/papers/webwriting/> [20th February 2014]
23. Peramunetilleke, D, Wong, R (2002) *Currency exchange rate forecasting from news headlines*, *Australian Computer Science Communications*, 24(2), pp.131-139
24. Redish, J (2012) *Letting go of the words*, eBook, [Waltham, MA] : Morgan Kaufmann

25. Rogers, T (2014) *Newswriting: Use verbs & adjectives wisely* [Online] Available:
<http://journalism.about.com/od/writing/a/verbsadjectives.htm> [20th February 2014]
26. Rogers, T (2014) *Subject-Verb-Object Exercises* [Online] Available:
<http://journalism.about.com/od/writing/a/Subject-Verb-Object-Exercises.htm> [20th February 2014] / <http://journalism.about.com/od/writing/a/Use-Subject-Verb-Object-In-Your-Writing.htm> [20th February 2014]
27. Rogers, T (2014) *Write Great Headlines* [Online] Available:
<http://journalism.about.com/od/writing/a/headlines.htm> [20th February 2014]
28. Ross, S. (2001) *Writing for the Web* [Online] Available from:
<http://www.selectideas.co.uk/ross/jcareers.htm> [20th February 2014]
29. Sperber, D, Wilson, D (2004) *Relevance Theory*, in G. Ward and L. Horn (eds) *Handbook of Pragmatics*, Oxford, Blackwell, pp. 607-632
30. Turner, D, W (2010) *Qualitative Interview Design: A practical guide for novice investigators*, *The Qualitative Report*, 15(3), pp.754-760
31. Website Optimization (2009) *Interview with Peter Pirolli* [Online] Available from:
<http://www.websiteoptimization.com/speed/tweak/information-foraging/interview.html>
[20th February 2014]
32. Winstanley, C (2014) *Dissertation Research Interview Designs* [online] available:
<http://www.dummies.com/how-to/content/dissertation-research-interview-designs.html>
[2nd July 2014]

APPENDICES

Appendix A. SURVEYMONKEY SURVEY DESIGN

Main page design and an example extract of survey questions, demonstrating the news excerpt and subsequent headline choices.

Thomson Reuters - News Headlines
Education

Summary Design Survey Collect Responses Analyze Results

Edit Survey Preview Survey Send Survey

Survey Options

Question Bank ?

Print Survey

Restore Questions

Page Randomization

Edit Survey

To change the look of your survey, select a theme below.

Aqua Create Custom Theme

TITLE & LOGO Edit Title Edit Logo Edit Layout



THOMSON REUTERS

Thomson Reuters - News Headlines

Q1 Edit Question Add Question Logic Move Copy Delete

1. (Reuters) - The U.S. Navy said on Tuesday it continued to examine options for a system to help Northrop Grumman Corp's new MQ-4 Triton high-altitude unmanned spy plane detect and avoid other aircraft, and expected to decide on a new approach in less than a year.

U.S. Navy considers options to ensure drones avoid other planes

U.S. Navy considers options to help new MQ-4 Triton spy drones avoid other planes

U.S. Navy mulls options to help drones avoid other planes

+ Add Question Split Page Here

Q2 Edit Question Add Question Logic Move Copy Delete

2. (Reuters) - A pioneering European Union survey into the impact of pests and diseases on honey bees found death rates were lower than feared, in part countering concerns about the collapse of colonies of the crop-pollinating insects.

European Union survey finds honey bees death rates lower than first thought

EU study finds honey bees death rates are lower than feared

Pioneering EU study finds honey bees death rates are lower than feared

Appendix B. SURVEYMONKEY QUESTIONS WITH BEST PRACTICE MIX

9th April 2014 (<http://feeds.reuters.com/reuters/scienceNews>)

Combo	Headline
C1C2, C3C4	U.S. Navy mulls options to help drones avoid other planes
C1C2,C3C4	EU study finds honey bees death rates are lower than feared
C1C2, C1, C2	Scientists regenerate immune organ in mice
C1C2, C1, C2	Paralyzed patients regain movement after spinal implant: study
C1, C2, C3, C4	U.S. Navy to test futuristic, super-fast gun at sea in 2016
C1, C2, C3, C4	Warm blooded or cold? Dinosaurs were somewhere in between
S1S2, S3S4	Sea creature fossil found with oldest-known cardiovascular system
S1S2, S3S4	Scientists dismiss claims that Yellowstone volcano about to erupt
S1S2, S1, S2	Europe launches satellite to monitor environment, disasters
S1S2, S1, S2	Birds of a feather: hummingbird family tree unveiled
S1, S2, S3, S4	Multinational crew blasts off, arrives at space station
S1, S2, S3, S4	Shaking Shakespeare: Richard III was no hunchback after all
F1F2	China says Jade Rabbit moon rover alive but weak
F1F2	Climate changes push native Montana trout toward extinction:study

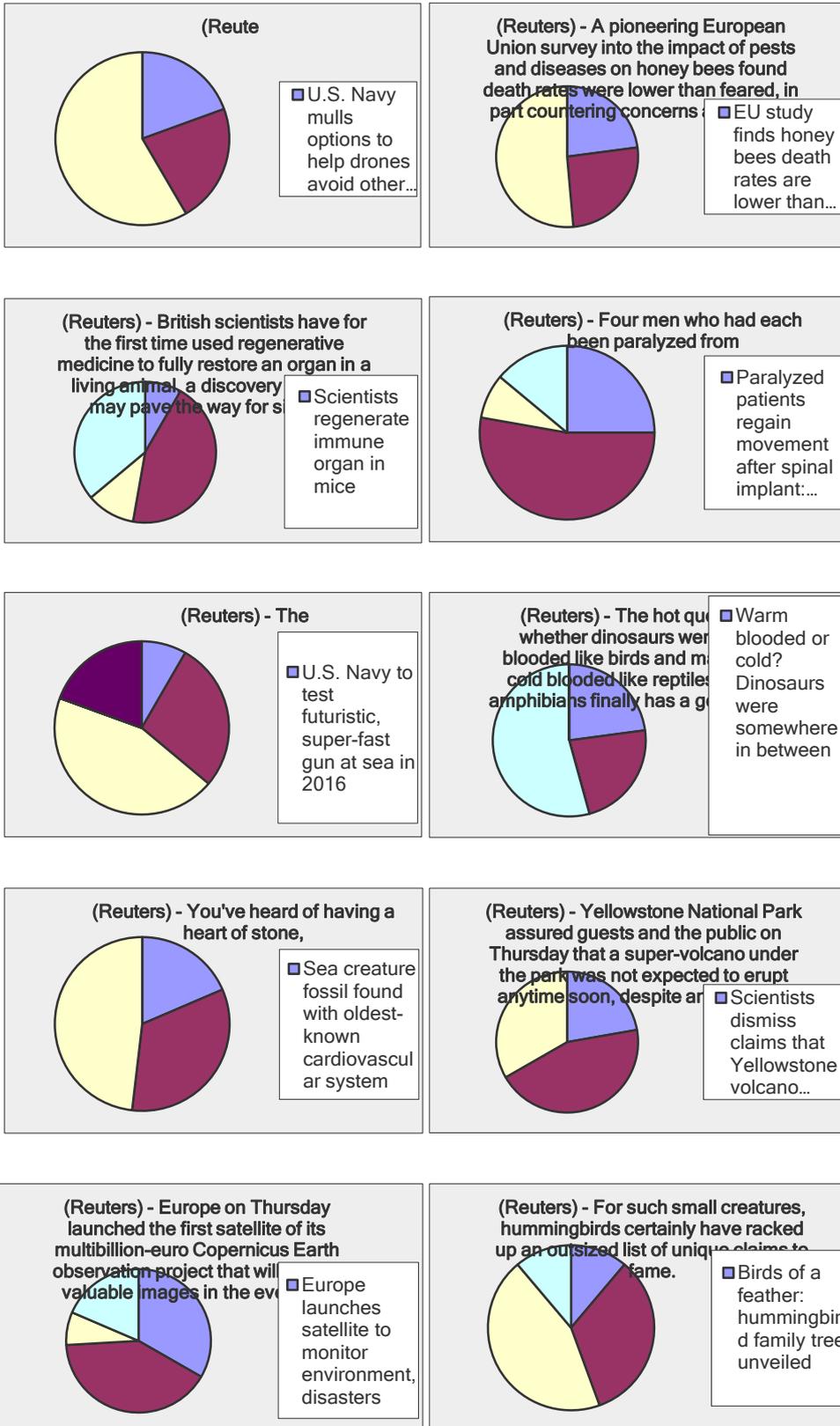
9th April 2014 (<http://feeds.reuters.com/reuters/globalmarketsNews>)

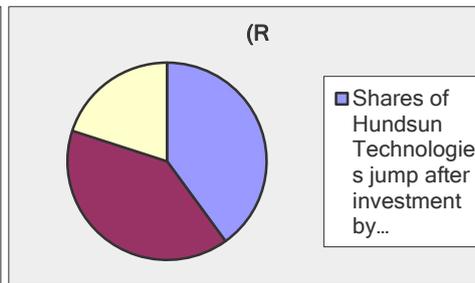
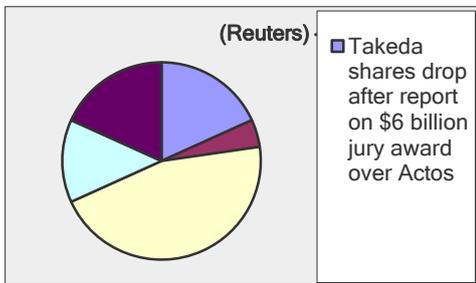
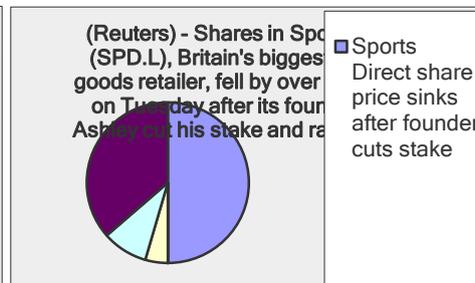
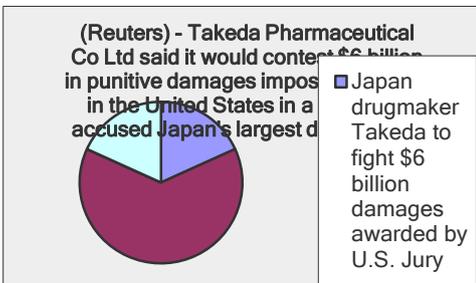
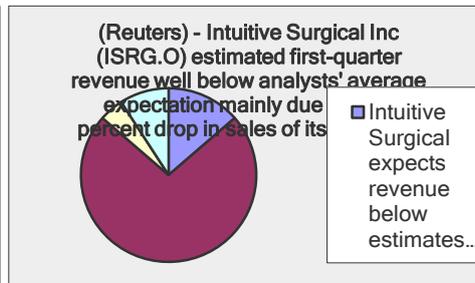
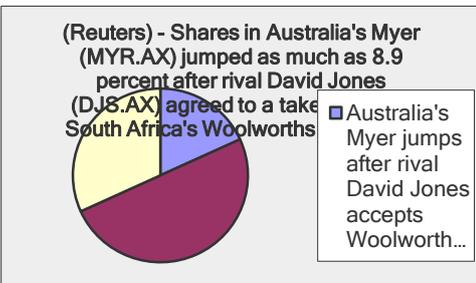
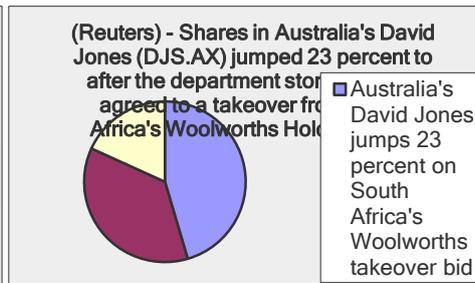
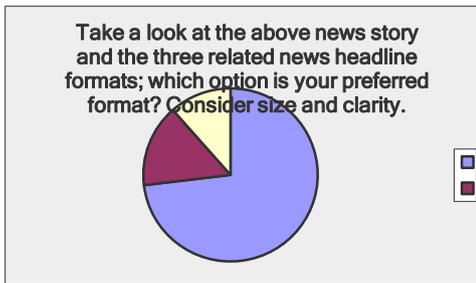
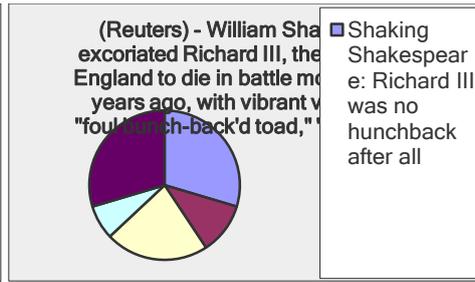
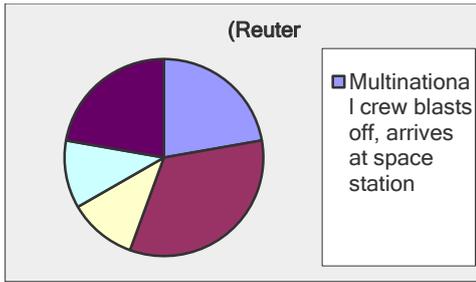
Combo	Headline
C1C2, C3C4	Australia's David Jones jumps 23 percent on South Africa's Woolworths takeover bid
C1C2,C3C4	Australia's Myer jumps after rival David Jones accepts Woolworths takeover
C1C2, C1, C2	Intuitive Surgical expects revenue below estimates as da Vinci sales fall
C1C2, C1, C2	Japan drugmaker Takeda to fight \$6 billion damages awarded by U.S. Jury
C1, C2, C3, C4	Sports Direct share price sinks after founder cuts stake
C1, C2, C3, C4	Takeda shares drop after report on \$6 billion jury award over Actos
S1S2, S3S4	Shares of Hundsun Technologies jump after investment by Alibaba's Jack Ma
S1S2, S3S4	HTC shares fall after bigger-than-expected first quarter loss
S1S2, S1, S2	Gigamon estimates first-quarter revenue below Street, shares slump
S1S2, S1, S2	Son of senior Shi'ite cleric sentenced to death in Saudi Arabia
S1, S2, S3, S4	Father blames government 'idiots' as California town mourns killings
S1, S2, S3, S4	Ukraine says fires warning shots at Russian helicopters
F1F2	Ukraine's empty seat at table darkens party for Putin's new ex-Soviet bloc
F1F2	China's state-owned sector told to cut ties with U.S. consulting firms

Appendix C. SURVEYMONKEY RESULTS, % PREFERENCE

Content Test 1					
	Original	c1c2	c3c4		
Scientific	19.4%	22.2%	58.3%		
Scientific	22.9%	25.7%	51.4%		
Financial	45.5%	36.4%	18.2%		
Financial	18.2%	50.0%	31.8%		
Content Test 2					
	Original	c1c2	c1	c2	
Scientific	8.3%	44.4%	11.1%	36.1%	
Scientific	25.0%	52.8%	8.3%	13.9%	
Financial	13.6%	72.7%	4.5%	9.1%	
Financial	18.2%	63.6%	0.0%	18.2%	
Content Test 3					
	Original	c1	c2	c3	c4
Scientific	8.3%	27.8%	44.4%	0.0%	19.4%
Scientific	22.9%	22.9%	0.0%	54.3%	0.0%
Financial	50.0%	0.0%	4.5%	9.1%	36.4%
Financial	18.2%	4.5%	45.5%	13.6%	18.2%
Structure Test 1					
	Original	s1s2	s3s4		
Scientific	18.5%	33.3%	48.1%		
Scientific	22.2%	44.4%	33.3%		
Financial	40.0%	40.0%	20.0%		
Financial	55.0%	35.0%	10.0%		
Structure Test 2					
	Original	s1s2	s1	s2	
Scientific	33.3%	40.7%	7.4%	18.5%	
Scientific	11.1%	33.3%	44.4%	11.1%	
Financial	26.3%	15.8%	31.6%	26.3%	
Financial	68.4%	10.5%	15.8%	5.3%	
Structure Test 3					
	Original	s1	s2	s3	s4
Scientific	22.2%	33.3%	11.1%	11.1%	22.2%
Scientific	29.6%	11.1%	22.2%	7.4%	29.6%
Financial	31.6%	5.3%	47.4%	15.8%	0.0%
Financial	15.8%	36.8%	21.1%	21.1%	5.3%
Format Test 1					
	f1f2	f1	f2		
Scientific	73.1%	15.4%	11.5%		
Financial	61.9%	23.8%	14.3%		

Appendix D. SURVEYMONKEY RESULTS, % PREFERENCE (VISUAL)





(Reuters) - Shares of HTC (2498.TW) opened down 0.6 percent on Tuesday after the struggling smartphone maker posted a worse-than-expected loss in the first quarter.

■ HTC shares fall after bigger-than-expected first quarter...

(Reuters) - Gigamon Inc (GIMO.N), which makes traffic management software, estimated lower-than-expected first-quarter revenue below Street, shares slump

■ Gigamon estimates first-quarter revenue below Street, shares slump

(Reuters) - A Saudi court on Monday sentenced to death the son of a senior Shi'ite Muslim cleric after he was convicted of shooting at forces in the kingdom's

■ Son of senior Shi'ite cleric sentenced to death in Saudi Arabia

(R) Father blames government 'idiots' as California town mourns killings

■ Father blames government 'idiots' as California town mourns killings

(Reuters) - A Ukrainian anti-aircraft battery fired warning shots to prevent an incursion on Saturday to help prevent gunships from the peninsula, the Ukrainian

■ Ukraine says fires warning shots at Russian helicopters

Take a look at the above news story and the three related news headline formats; which option is your preferred format? Consider size and clarity.

■ [Small square] [Medium square] [Large square]

Appendix E. DS PROPOSAL

Computing Project Proposal

Version V3

Student's Name: Chris Blatchford

Student's Number: H00019894

Student's Email Address: chris.blatchford@my.ohcampus.com / cblatcho@gmail.com

Project Title: Writing for the web: Developing a framework to determine efficient communication of key data

Proposal Submission Date: 7th January 2014

V2 proposal Date: 12th January 2014

V3 proposal Date: 6th February 2014

Version Number of the Proposal: v3.0

DA Class ID: UKL1.CKIT.702.H0028503

Name of DA: Paul Kingsley

RMT Class ID: LAUR-906-201424-2

Name of GDI: Bendick Mahleko

Ethical Checklist Completed: Yes

Name of SSM: Mila Zimbarno

The Program: MSc in IT

Domain: CKIT-503-2

Proposal approved by: *(To be filled by the Faculty Manager)*

Date of the approval: *(To be filled by the Faculty Manager)*

Verification of the Proposal's Domain: *(To be checked and verified by the Faculty Manager – a list of modules)*

Sponsor's Details: NA

Sponsor's Background: NA

Sponsor's Agreement: NA

The Project Aims and Objectives:

The delivery method and subsequent structure of news via the web varies considerably depending on the content being delivered; financial information needs to be fast, reliable and succinct but also verifiable and sanitised; world news needs to be delivered faster than competitors (in a commercial environment), provide valuable insights and contain content that is unique to the provider. News from other areas of industry such as Intellectual Property, science, tax/accounting, legal and many others have their own specific requirements for news delivery.

There is existing research which suggests web pages are consumed and ingested differently to traditional paper based content, which has resulted in a number of guidelines proposed for writing for the web. For example, Nielsen suggests that scannability, Navigation and specific highlighting can increase the readability of digital text. A key component of online readability is the increased use of mobile devices, leading to limited screen entity to delivery information; all of the rules Nielsen outlines still apply, however it is my belief that this research can be extended further through analysis of specific types of information contained within a headline or news snippet. For example, the below provides sample headlines broken down into suggested categories :-

European shares broadly steady; IBEX hit by LatAm rout

European	shares	broadly steady	IBEX	hit by	LatAm	rout
Location	Asset	Behaviour	Exchange	Action	Location	Behaviour

European shares inch lower after soft Chinese data

European	shares	Inch lower	After Soft	Chinese	Data
Location	Asset	Behaviour	Action	Location	Data

Business morale, Bundesbank point to stronger German growth

Business morale	Bundesbank	Point to	Stronger	German	Growth
Behaviour	Entity	Action	Action	Location	Behaviour

UK considers pensions shake-up with uncertain results

UK	Considers	Pensions	Shake up	With uncertain results
Location	Behaviour	Asset	Action	Effect

A set of hypotheses will be developed during the specification and design phase; one such example hypotheses is that specific ordering of information can be more efficient to the human user when ordered in a specific manner. The overall aim of the project will be test a variety of rules and guidelines against several different types of news information through the construction of a

recommended framework, which provides best practice for authoring and structuring global news data.

Project Outline

The project will run through four main stages of production:

5. Firstly, a review of existing literature and research papers on web authoring (including traditional literary reviews of writing best practice), understanding how the web/internet changes this dynamic and how writing for the web has transitioned over the past decade, resulting in our initial hypotheses.

6. Secondly, design of the Project Specification & Design, detailing practical research into Thomson Reuters global and financial news delivery, with interviews and questionnaires designed to understand what structure and delivery mechanism are the most favoured, whether existing methodologies and frameworks are utilised, and what are the most effective mechanisms.

7. Thirdly, based upon the research results, we will develop a framework which will provide recommendations on how to approach the subject matter.

8. Fourthly, the framework will then be evaluated utilising real world operational data in order to fully test and validate the hypotheses.

Literature Survey / Resources' List:

The main themes running through the researched literature can be summarised under three main headings (for the purposes of this research paper); writing for the web (generic), digitizing news and effective news headline authorship. For example, Morkes & Nielsen (1998) suggests that "conciseness, scannability, and objectivity" are key tenants of writing for the web, resulting in "faster, more efficient processing of information". This is part of a more general "rule set" that can be applied to authoring content online, however news content changes that paradigm once again. Dor (2003) goes into some detail around news headline authorship, defining newspaper headlines as "relevance optimizers: They are designed to optimize the relevance of their stories for their readers.", which is a component part of the news headline structure, in addition to delivery of specific information in the minimal amount of space (especially for mobile device usage). Develotte and Rechniewski (2013) go into further detail around cultural and more detailed linguistic factors which may influence structure and interpretation; for example, they discuss perspective and the "role played by headlines in orienting the reader's interpretation of subsequent 'facts' contained in the article".

Scholarly Contributions of the Project

The scholarly objectives of the project are to contribute to the existing research and knowledge of best practice around writing for the web, by focusing on specific industry sector news content type. Writing for the web is more concerned with the end user objectives as opposed to traditional literature, especially with the increased commoditization of information on the web. Content needs to provide a) insights based upon exclusive content and capabilities and b) structure and delivery method.

The ultimate output is both academic and commercial in nature; academically, we want to better understand how specific usage of certain data/information varies across industry, and develop a methodology for structuring web information specific to that industry area, and commercially as a model for Thomson Reuters to potentially utilise for more efficient news delivery.

Description of the Deliverables:

Content of the project: Academic & industry research (Thomson Reuters), usability criteria, survey and interview questionnaires, statistical analysis of survey results. **Results deliverable:** a framework to help define structure of web content for the global and financial news sector. Recommended additional research would be development of a framework which takes into account key requirements/elements of various industry sectors for commercial entities.

Evaluation Criteria:

The primary evaluation criteria for the research will be the evaluation of the resulting framework; this will confirm whether the initial hypotheses developed during the design and specification phase are confirmed or rejected.

We will achieve this through the compilation of current day/week news stories (Source: Thomson Reuters) and then generate headlines through use of the developed framework. These headlines will then be matched to their real world counterparts, generated by Thomson Reuters, as well as other external news publishers (Economist, Bloomberg).

Both sets of data (framework generated and commercially generated) will be compared by:

- a) External parties (via social media) for comparison and ranking purposes; which news headline communicates “the best”
- b) Internal Thomson Reuters news editorial team;

Resource Plan:

Equipment: computer, news platform (TR), web server (testing of web page content), online survey service. **People:** Survey participants (TR), interviewees (TR)

Project Plan and Timing:

Stage of the dissertation writing process	Number of days/weeks needed	Start date	End date
Stage 1			
Initial reading (2013)	15	December 9 th	December 27 th
Stage 2			
DS Proposal & ethical sign-off	30	January 6 th	February 14 th
Stage 3			
Continued research (further development of a set of hypotheses)	25	February 14 th	March 21 st
Stage 4			

Project Specification & Design (detailed plan of the practical implementation and next steps of the research)	40	February 14 th	11 th April
Stage 5			
Practical field work (collection of data via survey's/questionnaires, development of a framework and evaluation of the same against our initial set of hypotheses).	30	11 th April	23 rd May
Stage 6			
Initial Writing		23 rd May	20 th June
Stage 7			
First draft		20 th June	4 th July
Stage 8			
Final draft		4 th July	18 th July
Submission			

Risk Assessment:

The biggest risk to this research/paper is the availability of the Thomson Reuters news platform team, specific individuals who are responsible for authoring, writing, converting and delivering the content to our end clients. One key element of how news content reaches the end user is the "chain of custody", or how the original format and content reaches the end of it's journey. Journalists in the field are the "receivers" of information directly; this is then converted, translated, reviewed (for structure and delivery) and delivered. It would be important to understand the various influencing factors in news content delivery, and access to these individuals (journalists, editors, publishers, translators) will likely be limited given their geographical locale, along with their time constraints.

Quality Assurance:

Regular check-in with the assigned DA (Paul Kingsley), responding to feedback in a timely manner, with staggered milestone goals according to the above scheduled "STAGES" to check on progress.