Factful: Engaging Taxpayers in the Public Discussion of a Government Budget

Juho Kim[1] **Eun-Young Ko**[2] **Jonghyuk Jung**[2] **Chang Won Lee**[2] **Nam Wook Kim**[3] **Jihee Kim**[2]
[1] MIT CSAIL, Cambridge, MA USA
[2] KAIST, Daejeon, Republic of Korea
[3] Harvard SEAS, Cambridge, MA USA
juhokim@mit.edu {key44, heretor92, chiyah, jiheekim}@kaist.ac.kr namwkim@seas.harvard.edu

ABSTRACT
While a government budget determines how taxpayers' money is allocated to various programs and stakeholders that compete for limited resources, the extensiveness and complexity of the budget and its process hinder taxpayers from understanding the budget information and participating in the public discussion. To engage taxpayers in the public discussion around budgetary issues, we leverage news articles containing budgetary information for design opportunities. We present Factful, a web-based annotative article reading interface that enhances the article with fact-checking support and contextual budgetary information by processing open government data. In our lab study, participants using Factful discussed more critically with more fact-based supporting statements. They built a rich context surrounding the relevant budget facts beyond what was presented in the article. Factful presents a simple yet powerful model for supporting fact-oriented budgetary discussions online by leveraging open government data.

Author Keywords
Budget; Discussion support; Open government data; Annotation tool; Fact-checking; Deliberative democracy.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation

INTRODUCTION
The idea of "deliberative democracy" emphasizes that citizens should actively participate in decision making and related discussions [6]. Open and public discussions are essential to realize its ideal, whose characteristics include informed (i.e., arguments should be presented with accurate and relevant information), balanced (i.e., conflicting views need to be considered), and substantive (i.e., arguments should be judged by evidence and not by other external factors), among others [8]. As more and more public discourses take place online via social media, news outlets, and discussion forums, there is a growing attention to support a deliberative democracy in online discussion environments. The HCI community has actively analyzed existing online discourses and designed systems to promote informed and balanced discussions [4, 18, 19, 21, 25]. On the other hand, with more governments engaging in open data policies, there are opportunities in using data to better inform online deliberation. In this work, we specifically focus on supporting **budget-related discussions online** by leveraging open budget data.

A government budget is a layout of the government's financial plans on how to spend its revenues (e.g., tax), which is its answer to how to best allocate limited resources. It contains an immense amount of information divided into a large number of programs and services, as everyone's life in a constituency is involved. Moreover, the budget and its processes are complex and multidimensional. There are numerous decisions to make in budget procedures, which determine how budgets are prepared, approved, and carried out. In making the decisions, various interests and tradeoffs must be considered. Although such complex and subtle nature of budget lends itself to public discussions [11], most taxpayers do not actively engage in them. Firstly, the extensiveness and complexity of the budget and its processes hinder taxpayers from understanding how their tax money is spent and which government programs they can benefit from. Secondly, not many interactive channels exist for taxpayers to participate in discussions, as well as stay informed and engaged.

This research addresses the question of how to engage taxpayers in the public discussion of a government budget online. To identify the challenges taxpayers face in accessing and understanding budgetary information and participating in budgetary discussions, we conducted a) 182-person online survey on budget awareness, b) interviews with three experts in a government budget sector, and c) interviews with five taxpayers. Findings reveal that the budget is extremely complex and extensive, making it difficult for taxpayers to understand it with a balanced view. Discussions around budgetary issues are often intertwined with objective facts (e.g., amounts allocated to programs) and subjective opinions (e.g., stakeholders competing against each other for more resources). Such complexity makes it difficult to engage in more meaningful discussions for experts and taxpayers alike.

For many taxpayers, media outlets were the primary channel through which they learned about the budget-related issues.

CHI 2015, April 18–23 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04$15.00
http://dx.doi.org/10.1145/2702123.2702352

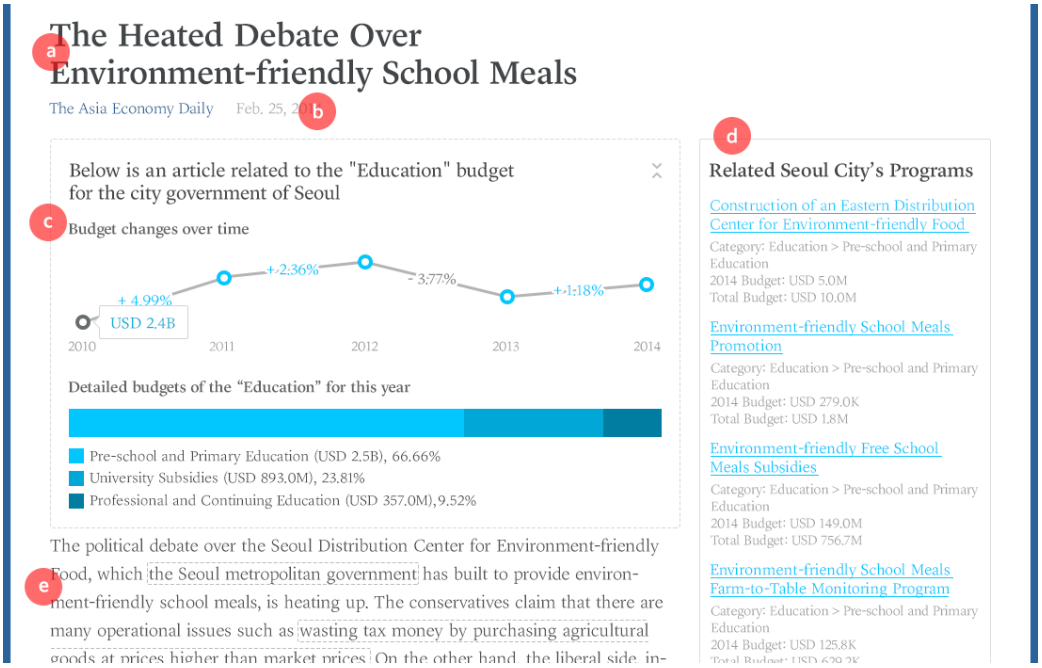

Figure 1. Overview of Factful powered by open budgetary data from the Seoul metropolitan government: (a) title of the article, (b) posted date and news outlet, (c) contextual budget category panel (education), (d) individual budget programs relevant to the article, (e) article content.

We also noted that taxpayers are conscious of how fact-based and balanced the article is, and that they often rely on social channels such as reader comments, social media, and discussion forums to get additional information. We conclude that there are unique opportunities to raise public awareness and support discussions on budgetary issues by leveraging the article reading experience, which is already familiar and social to many taxpayers. Specifically, we explore two ideas: 1) provide customized support for **fact-checking** and 2) embed **contextual budget information** related to an article.

To support fact-oriented discussions around budgetary issues, we built Factful, a web-based interface for reading articles (Figure 1). A reader can request fact-checking on any part of the article through in-place annotations. Other readers reading the same article can respond to the request by adding relevant references and scoring their confidence in the references. Factful also automatically analyzes the article text and inserts related budget information, including semantically matching programs, temporal changes in budget size, and programs of similar budget sizes. Factful also supports annotation-based threaded comments for general discussion.

To evaluate the effect of fact-checking support and contextual information in promoting budgetary discussions, we conducted a laboratory study where participants read budget-related articles with discussion only, discussion + fact-checking, and discussion + fact-checking + contextual information. In our lab study, participants using Factful discussed more critically with fact-based statements. They built a rich context surrounding the relevant budget facts, beyond the article content. With a more critical view and richer context around the budget-related articles, they engaged in fact-oriented discussions more actively.

The contributions of the paper are as follows:

- A survey and interviews with budget experts and taxpayers that revealed design challenges for improving public awareness on budgetary issues.
- Factful, a web-based article reading application with annotation support for fact-checking and automated support for contextual budget information.
- Results from a laboratory study showing that participants using Factful showed more critical view on the article and discussed with richer context.

RELATED WORK

We discuss previous work on supporting political discussions online, with specific focus on fact-checking and reading support with contextual information.

Political Fact-checking

Fact-checking normally refers to the act of verifying the truth of public facts and claims. Political fact-checking is most common, which is normally done by journalists who specialize in assessing political claims by reviewing large amounts of documents [13]. Media companies or non-government organizations publicly share their fact-checking results online, from websites such as `factcheck.org` and `politifact.org`. To overcome the labor-intensive and costly nature of expert-driven approaches, automated or community-driven solutions to detect and even fix misinformation have been introduced. Computational fact-checking [29] models claims into parameterized queries to fight vague or questionable statements. Finder [7] combines user-reported and automatically mined disputable claims to highlight disputable claims on any web

page. Reddit had a successful experiment in community-powered political fact-checking during the U.S. presidential debate [1], in which anyone could submit a political claim for the community to verify. On-demand fact-checking has been explored in the crowdsourced voters' guide [17]. Maater [22] attempts to tap into the wisdom of the crowd for fact-checking, which lets readers contribute information sources and improved assessments. Our work extends these crowd-moderated approaches to provide fact-checking specifically for the domain of government budgets.

Involving taxpayers in the fact-checking process may also align with how people accept new facts. Previous research argues that learning a new political fact and correcting misinformation is not a simplistic process [10], which involves people to internally assess new information against their current beliefs [30]. Accepting this view, fact-checking should be seen a persuasive process rather than a point solution of showing or checking facts. Meola et al. [23] and Metzger [24] argue that fact-checking approaches should be catered to the ability and motivation of users, and advocate for techniques that allow responding to tasks that are most needed, comparing multiple references, and considering peer or editorial review. Garrett and Weeks [10] argue that real-time corrections to political misperceptions can raise resistance, and suggest socially recommended corrective messages as a potential solution to avoid biases. Factful attempts to enable social interactions around facts with annotation-based fact-checking requests and responses.

Supporting Political Discussions Online

Online news services and discussion forums, while they lower the barriers to participation for anyone interested, also face challenges in monitoring the quality of comments. Discussion forums implement different moderation mechanisms, such as community guidelines with dedicated moderators or aggregated crowd ratings [20, 26]. Research has shown that individual motivation in reading and writing online comments affects perceptions of quality [5].

While the web hosts diverse perspectives and information, selective exposure [9] prevents people from accessing them because individuals favor information that confirms and reinforces their current views, resulting in polarization. Previous research has introduced social mechanisms and interfaces to promote listening, balanced perspectives, and exploration of diverse opinions. Examples include visualizing the political lean in the reading behaviors [25] and source expertise indicators [21]. ConsiderIt [18] facilitates a fact-grounded political discourse with personal pro/con lists for an issue. Added social layers help users browse others' lists. Reflect [19] encourages active listening and deliberation in online comments. It adds a listening box to the comment section, asking users to summarize the points made in the original comment. Inspired by these systems, we innovate on budgetary discussion support that little previous research has focused on.

[1] `http://www.reddit.com/r/politicalfactchecking/comments/10w86s/official_politicalfactchecking_thread_for_the_9pm/`

Reading Enhancements

Enhancing the news reading experience with additional information is another method for promoting a healthier discussion. Anchored discussion is an alternative form of commenting that allows readers to directly annotate on specific parts of text (e.g., `medium.com`). The Reader's Helper [12] detects the reader's topics of interest and automatically annotate phrases containing these topics to help readers quickly locate items of interest. Tell Me More [15] automatically fetches stories that provide additional yet relevant quotes, actors, figures, and information and presents snippets in the sidebar of the screen. Videolyzer [4] supports credibility evaluation of claims via video and text annotations. Contextifier [14] augments articles about a company with annotated stock visualizations. Kong et al. [16] applied crowdsourcing to extract references between text and charts to maximize the effect of visual reinforcements in charts. In an educational setting, an interface supporting anchored annotation led to more frequent and specific comments than a discussion board [1]. Factful takes a similar approach to these systems by analyzing the article text and inserting contextual information to relevant parts.

SURVEY & INTERVIEWS: KEY OBSERVATIONS

To better understand the current challenges taxpayers face in accessing, processing, and discussing with budgetary information, we conducted a web-based survey as well as a series of interviews with both taxpayers and experts.

Taxpayer Survey & Interviews

We designed a web-based survey to gauge taxpayers' budgetary awareness in South Korea. The survey asked respondents about their perception of the government budget, followed by quiz questions that asked to estimate the size of various budget programs. We also inserted a simple A/B test for a question that asked people's opinions about whether to increase or decrease the roadwork budget. In one condition, we displayed detailed contextual information about related programs and their relative sizes against the roadwork program. The other programs were sibling programs in the transportation and logistics category, which is the parent category of the roadwork program. The goal of this setup was to see if presenting contextual information affects people's opinions.

Responses were collected by an external commercial survey agency that controlled for the age, income level, and political affiliations of the respondents. Total of 182 people responded (82 female), with our target age (20 - 59) reflecting the age distribution of the Korean population as of 2013: 20s (21%), 30s (25%), 40s (31%), and 50s (23%). 82% had college or higher education. Self-reported annual income ranges showed 10% with no income, 66% below USD 46K, and 24% higher than USD 46K (Korea's GDP per capita in 2013: USD 26K). Self-reported political affiliations showed 41% liberal, 30% conservative, 18% neither, and 11% not specified.

To complement the survey data with a qualitative understanding of taxpayers' awareness and challenges, we conducted semi-structured interviews with five taxpayers: three college students with ages 20, 22, and 23 each, and two income-taxpayers, one in the top 20% income percentile and the other

in the top 1%. Each interview session took about 30 mins. The interviewees were guided to articulate their knowledge and experiences on a government budget. Here we summarize main findings from both the survey and the interviews.

1. Low awareness and interest in budgetary issues. In 7-point Likert scale questions (1-negative, 7-positive), survey respondents showed moderate interest in knowing about the government budget (4.6) and believed knowing about it is useful to them (4.5), but indicated they were not quite knowledgeable (3.1) and didn't actively search for budgetary information (3.2). On a question about estimating the relative proportion of the military budget within the entire government budget (10.2% as of 2013), respondents' estimates were almost twice higher (20.26%, median=20, SD=14.05). When we asked survey respondents to select a budgetary program larger in size between two options (e.g., welfare vs military), only 39% of the answers were correct, averaged over four such comparison questions per session. While incorrectly guessing simple numbers may not directly correspond to lack of awareness, many respondents expressed interest in improving their budgetary knowledge.

Interviewees felt disconnected from budgetary information, as an interviewee said, *"I feel distant from all the big numbers that don't really mean anything to me."* They also expressed frustration that there is no channel to make their voices heard in the process of writing a budget proposal, executing the budget, or settling the budget. Echoing the survey results, interviewees commented that they didn't have enough budgetary knowledge, which in turn made them reluctant to participate in online discussions. As an individual's influence is rather insignificant, taxpayers responded that their interest and awareness on budget-related issues are short-lived and volatile, unless anything directly concerns their daily lives.

2. Contextual information matters in opinion formation. In forming opinions about budgetary issues, access to relevant and contextual information may matter. In the A/B test setting that asked survey respondents to indicate their opinion on the change of the roadwork budget, the group with no contextual information replied that it should increase by 22.6% on average, while the group with contextual information only desired a 9.2% increase ($p = 0.01$ with Mann-Whitney's test, $z = -2.5$). No significant differences were found between different age, income, or political affiliation groups. Where does this difference come from? A possible explanation is that by seeing other programs in the same category, respondents might have realized the tradeoff involved in raising the roadwork budget and compared the relative importance against other programs that were displayed. This can be especially true for a majority of taxpayers with limited awareness and understanding of the budget, who are less likely to hold a strong current view that may add resistance. While this simple intervention is not meant to conclusively demonstrate the value of contextual information, it provides an insight that contextual information may affect how people form opinions about budgetary issues.

3. News media are the familiar and preferred way to learn about budgetary issues. A majority of our survey respondents said they regularly read articles online (74%), with 7% via social networks. This is comparable to the U.S. data, which shows 50% of population reading online news and 10% on social networks in 2012 [2]. Interviewees also confirmed that news outlets were their primary source of information to learn about budgetary issues. None had attempted to read government reports, and they found official reports to be hard to access and interpret when shown by the researchers. The interviewees mentioned that information presented in news articles is more accessible because they are written in a comprehensible, engaging manner. They were also well-aware of the fact that information presented by media is potentially biased. They also pointed out that other people's comments help them recognize the potential subjectivity, bias, or error in an article. Comment sections to news articles serve as an alternative source of learning through a means of social interaction. A U.S. survey aligns with this observation, with 37% of users finding commenting on news stories as an important feature to have, and 25% have a commenting experience [27].

Expert Interviews

We further conducted semi-structured interviews with three experts in the government budget domain: a former congressman, a current staff member to a congressman, and a current government official in the Ministry of Strategy and Finance in South Korea. By interviewing domain experts, we hoped to understand why channels for accessing and discussing government budgets are not very public-facing, and what challenges existing resources and media face in raising taxpayers' budgetary awareness. Each interview session took about 1-2 hours. The interviewees were asked to walk through the budget processes they had worked on, and to elaborate on how they made budget-related decisions and handled the complexity of the budget. Here we summarize key findings drawn from the interviews.

1. Budget-related issues are multifaceted. All three interviewees emphasized that the budget entails various goals such as economic growth, job creation, and solid social safety net, so that careful decision making and discussions with multiple interest groups and government ministries are needed.

2. Budget is too extensive for a small group of experts. We observed that the budget procedures cover extensive information to be handled by a small number of experts in the executive and the legislature. Therefore, fact-checking itself can become a big challenge and sometimes important matters can be missed in the process even among experts. An interviewee said, *"The system is short-staffed ... it's impossible to check every detail of a government budget in a review process."*

3. Budget-related processes are black boxes to taxpayers. We observed that budget-related facts and processes are often obscure and not accessible to taxpayers. An interviewee said, *"(Taxpayers) sometimes only see their own interests and fail to realize that compromises need to be made. Therefore, it is challenging for us to accommodate what taxpayers want."* Moreover, we observed that publicly available resources are often presented poorly for taxpayers to develop the necessary perspectives. Budget reports are often dispersed across mul-

tiple websites of ministries and agencies, leaving the public in disarray when they try to gather information.

DESIGN GOALS

Our findings from the survey and interviews imply that budgets are complex, contextual, and extensive, and that taxpayers commonly rely on news articles to access budget-related information. To addresses these challenges and promote open budgetary discussions, we decided to augment the article reading experience already common to many, rather than design a completely new channel. We have identified design goals for a **budgetary discussion support tool** below. The individual design goals provide concrete guidelines to encourage public participation in budgetary discussions while reading news articles.

Present rich, contextual budgetary facts while reading. The taxpayers in our interviews listed the lack of relevant knowledge as one of major factors which discourage them to engage in a discussion of a government budget. Accessing relevant facts and supporting evidence can help construct a richer context, potentially leading taxpayers to construct more informed opinions. A reading interface can fetch relevant facts for easier access, thus reducing the need to search for external data. As a result, discussions can also be more fact-oriented and focused on budgetary issues.

Provide built-in methods for evaluating the credibility of facts and claims. The taxpayers in our interviews raised questions on the objectivity of news media. They expressed concerns around lack of diversity in the political spectrum certain media cover and potential biases in information sources and interpretations. Social- or crowd-based moderation can be a potential solution, as interviewees suggested that they rely on other's comments and statements to complement their understanding. Interface support can be designed to rate the credibility of information sources, compare multiple possible sources for a fact, and vote on other's facts and claims. Embedding the fact-checking process and discussion into an article itself can potentially lower the barrier to participation.

Add structured, lightweight, and multiple ways to participate. According to our survey and interviews, a common challenge taxpayers faced in joining budgetary discussions was lack of confidence in their ability to contribute. Therefore, it would make sense to design multiple ways to participate, especially lightweight and structured methods that require little effort and domain understanding, to motivate users who are not knowledgeable about the matter. We explore raising annotated questions and requesting fact-checking as lightweight methods. Having multiple ways to contribute allows users to engage at different depths, depending on personal interests or level of commitment.

Let readers initiate discussion with meaningful chunks of information. Due to the complexity of budget discussions, users can benefit from being able to scope and initiate discussion with any level of information they are interested in. Supporting annotative discussion can give users more control in what the discussion should be about. In most cases, budget-related conversions pertain to a specific topic, while official documents tend to hold more information than the topic covers. The implications are that the information needs to be refined and broken down into digestible units for readers to encourage participation. Annotations at a finer-grain level can reduce the cognitive load for readers and promote a clearly scoped discussion.

FACTFUL: BUDGET-AWARE READING INTERFACE

This section introduces **Factful** (Figure 1), a system for supporting budgetary discussions online by enhancing the article reading experience. Factful analyzes an article, queries the budget database to fetch relevant budgetary facts, and adds the fact-checking and contextual budget information layers to the article reading UI. While the **contextual budget information** layer presents automated, ambient information that helps readers learn about relevant facts, the **fact-checking** layer supports active, social discussions that are fact-driven. We expect the two layers to complement each other in serving the goal of improving budgetary awareness and discussion.

Processing Budget Data

To serve snippets of budgetary facts at various granularities (e.g., article-, number-, and word-level), Factful maintains a database of budgetary information per each budget program, indexed by its name, hierarchical category structure, and temporal changes in the amount. Our current database is built with budget data from the Seoul metropolitan government in South Korea. The original interface was designed in Korean and was translated to English for the screenshots in this paper. With the recent commitment to open government data, the Seoul metropolitan government was the first in the world to publicly release their internal official documents in April, 2014. They publish 75.9% of their internally communicated documents since 2013. Related to the budget, the published documents contain detailed information about daily budget transactions of each program and of each department.

To meet our specific needs to construct per-program references, we merged two datasets from the city's open government project [`opengov.seoul.go.kr`]. First, Factful processes and indexes these internal documents to construct a database for enhancing the reading interface. The second source is *Clean Budget*, which has the general budget information of each program, including this year's assigned budget, total budget, executed budget, etc. Our data processing module retrieves the budget category information of each program by mining the city's open government gateway website and processing the city's official documents. Our current dataset covers the city's programs in full, from 2010 to 2014. For 2014, the city government assigned approximately USD 24 billion, with its structure containing 13 first-level categories and 4629 individual programs (as of September 2014, continuously growing over time). The database construction needs to be performed once initially, although incremental program updates to the database can be made anytime.

Displaying Contextual Budgetary Information

When a user opens an article using Factful, Factful analyzes the article text, issues a set of queries to the database for relevant facts, and displays the retrieved contextual snippets in

the UI. We explore three types of contextual budget information, which we describe below.

1. Annual budget changes and article-level category: Factful presents the reader with an overview of a high-level budget that's most relevant to the current article (e.g., education, welfare, environment), as can be seen in Figure 1c. The idea is to equip the reader with high-level budgetary information before reading the article. Upon detecting the closest budget category, Factful generates a graph of annual budget changes for the selected category with the amount and a percentage of difference against the previous year. Factful also shows a list of third-level categories (e.g., pre-school education, general management of university from Figure 1c), which belong to the root-level category (e.g., education). The budget amount for the current year and the relative proportion within the category are displayed.

Our category detection algorithm works as follows: 1) all the budget program names are parsed and stored in a bag-of-words model; 2) frequent words within each category (above a pre-configured threshold), stop words, and short words (shorter than two characters, may be specific to the Korean language) are removed; 3) keep a tally of how many hits each category finds by going through each word from the article; 4) assign a "fit" score for the current category by summing all hits; and 5) pick the category with the highest score.

2. List of most relevant programs: When only presented with high-level budget category information, our pilot users complained that they would benefit more from concrete programs they could intuitively understand the purpose of. We decided to display a list of most relevant individual programs to the article (Figure 1d). Located in the sidebar, each entry displays the name, amount, higher-level categories it belongs to, and a URL that links to a government-maintained webpage with more information.

Our program suggestion algorithm picks top five programs for the current article, by using a customized TF-IDF algorithm [28]. Factful considers all program names as a corpus, and each program as a document. For each non-stop word in the article, the TF-IDF score is computed against each program. After all words are processed, the highest scoring N programs are presented. This simple TF-IDF method returned relevant programs for a number of articles spanning various domains in our internal tests. This may be the case because most budget-related articles focus on a specific topic and include many official budgetary terms. Some practical notes are that language-specific customization and weight adjustment with respect to word length (the longer the more weight, may be specific to Korean) affect results.

3. Automatic annotations with amount matching: Our interviews with taxpayers showed that they were turned away from a large amount of budget units that don't provide enough context in everyday lives. When monetary values are present in an article, Factful automatically highlights these values and annotates with budget programs that are of similar amounts (Figure 5b). The idea is to give readers comparison points, leading them to consider alternative options that the given

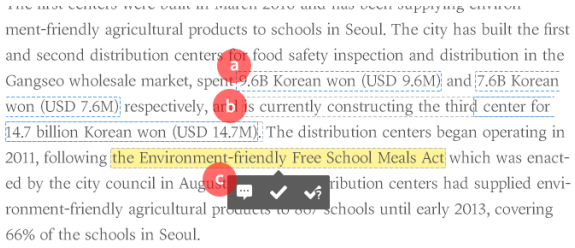

Figure 2. Dashed boxes indicate annotative regions for commenting, fact-checking and fact-check requesting: (a) a blue box indicates that Factful automatically annotated money amounts with budgetary information, (b) a gray box indicates there exists either users' comments, fact-checks, or fact-checking requests in the covered text, (c) a orange-highlighted indicates a current selection, which opens a popup menu with options for commenting, fact-checking, and fact-check requesting (from left to right). The menu opens when the user selects a dashed box or highlights content by mouse-dragging.

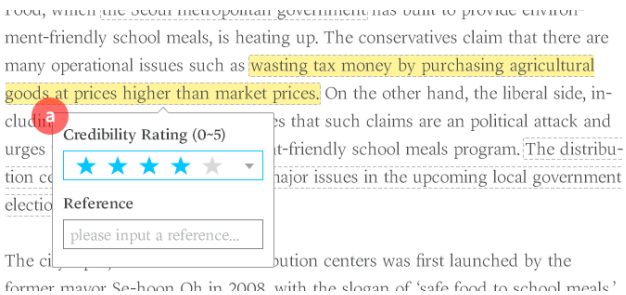

Figure 3. Annotative fact-checking interactions: clicking on the fact-check button opens a popup (a), so that a user can evaluate the credibility of the highlighted phrase in a 5-point scale, and add a URL reference that verifies the phrase.

amount in the article can achieve. Our amount-matching algorithm simply identifies all monetary values from an article with a rule-based detector, queries the database with items of similar amounts and categories, and displays top matches in the sidebar.

Annotative Fact-Checking

The fact-checking layer supports crowd-moderated, annotative fact-checking activities to encourage fact-driven discussions. Upon highlighting a phrase in an article, users can fact-check by adding a reference and assigning a credibility score, request a fact-check to make an open call, and start a threaded discussion with comments (Figure 2c).

Fact-checking: Factful has built-in support for fact-checking activities. Readers can initiate a new fact-checking activity by selecting a questionable phrase, or respond to a fact-checking request submitted by others. Fact-checking involves 1) providing an external reference to a statement made in the selected part of the article, and 2) evaluating how credible the statement is in a 5-point scale (Figure 3a). A fact-checked item provides a list of references and a brief summary of credibility ratings of the highlighted section (Figure 4a).

Fact-checking requests: Fact-checking requests are the most lightweight way to participate in the Factful-supported discussion. When a reader encounters a questionable statement, but does not have enough knowledge or feel committed to

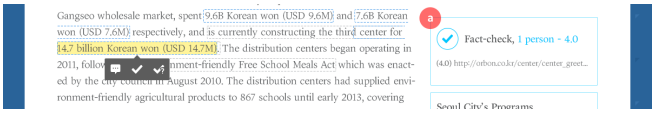

Figure 4. Fulfilling a fact-check request changes the sidebar display: (a) the number of completed fact-checks for the phrase, with a list of URLs entered and average credibility score.

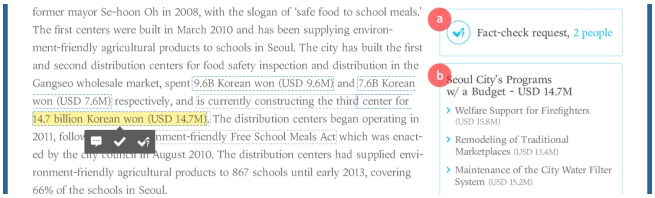

Figure 5. Budgetary fact enhancements for the selected phrase (orange-highlighted): (a) the number of fact-check requests for the phrase, (b) a list of Seoul city's programs most relevant and closest to the amount included in the phrase

actively engaging in the verification process, he can simply click the fact-checking request button. By separating the request and verification processes, tasks are smaller in size with low entry barriers. (Figure 5a). Despite the simplicity, fact-checking requests are still valuable because they initiate budgetary fact research and open discussions by leaving visual anchors for future readers.

Threaded discussion support: Factful supports threaded discussions based on annotations, as can be seen in other online venues like Medium [`medium.com`] (Figure 6). It is designed to serve as a place for open-ended discussions, complementing fact-checking activities and contextual budgetary information. Readers may leave a comment by highlighting a phrase and clicking the comment button (Figure 2c).

EVALUATION

To assess the effects of Factful on taxpayers as news consumers, we conducted a lab study that asked participants to read news articles on a government budget and discuss with others. We compared the baseline (annotative commenting only), Fact-checking (annotative commenting with the fact-checking layer), and Factful (fact-checking and contextual budgetary information) conditions. We hypothesized that reading news articles on budgetary issues with Factful leads to richer discussions. Specifically:

- H1. Readers will hold a more critical view on the article with Factful.
- H2. Readers will discuss with more fact-based statements with Factful.
- H3. Readers will build a rich context surrounding the relevant budget facts using Factful.
- H4. Readers will initiate and engage in the discussion with more existing discussion activities.

Participants

We recruited 38 participants from a behavioral study participant pool at a university. We divided them into three groups of 12, 13, and 13: the baseline, Fact-checking, and Factful groups, respectively.

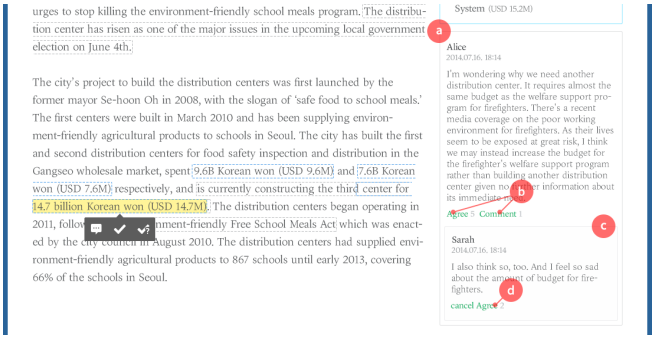

Figure 6. Threaded discussion support: (a) annotative discussion thread, (b) readers can agree (analogous to "like") or reply to an existing comment, (c) a threaded comment, (d) the cancel button for "agree".

Tasks and Procedures

Our lab study used a between-subject design (baseline, Fact-checking, and Factful), where the interfaces progressively included the additional fact-checking and contextual budgetary information support. The baseline condition supported threaded discussions, the Fact-checking condition supported fact-checking in addition to the baseline condition, and the Factful condition had contextual budget information in addition to the Fact-checking condition. In an hour-long session, each participant read three articles and engaged in unstructured discussions with a given interface, for seven minutes each. They were paid $10 for their participation and time.

We selected three articles that directly covered Seoul's budget and policies: 1) a political dispute regarding eco-friendly school meals, 2) a critical view on the Seoul metropolitan government for wasting its budget on advertising its tap water 'Arisu', and 3) a grand-scale waterway construction project. Participants could see activities of others who used the same interface as they did, and engage in discussions with them. We ensured that each participant read articles in a randomized order, meaning that they will see different levels of activity (i.e., the first article they read is likely to have much less discussion than the third one).

For each article, participants answered questionnaires before and after reading an article. The pre-reading survey asked about self-evaluated interest and informedness on the subject matter of the article in 7-point Likert scale (1-low, 7-high), while the post-reading survey included the questions about their interest and informedness again, and how credible and fact-based the article was. In addition to reading articles and filling out surveys, participants were allowed to search any relevant information on the web. The session ended with a final survey about the experience of each conditions' features and their usability.

Methods for Analyzing Discussion

To compare the quality of discussion we conducted a discourse analysis and quality assessment. To understand the characteristics of each comment, we coded the comments based on two dimensions: the topic that a comment addresses (related to budget, policy, or else), and the intention of a comment (express an opinion, provide information, or ask a ques-

1. Topic: Budget, Policy, Others

2. Intention

Opinion	w/ objective basis
	w/ subjective basis
	w/ no basis
Information	new & verifiable
	confirming existing
	unverifiable
Question /Info request	objective
	subjective
Simple statement	

Table 1. Discussion coding categories: each comment is coded with one of the 25 categories.

tion). Chen et al. [3] have explored similar fact attributes and comment categorization. The intention category was further segmented as shown in Table 1. This coding scheme led to total 25 exclusive categories (3x8 = 24 + simple statement such as "nice"). Two researchers extensively discussed the coding scheme with a few examples to arrive at the presented scheme. They then independently classified each comment into the 25 categories in a blind condition. Note that comments containing several sentences with different intentions were separated into multiple comments as they were heterogeneous in nature. For the total of 404 comments from all study sessions, inter-rater reliability was 0.613 (unweighted Cohen's Kappa). The two researchers discussed in person to reach agreement and finalize the labels.

Additionally, we asked five external raters to rate the quality of discussions using dimensions derived from Fishkin's work on deliberation [8]: informed, balanced, conscientious, substantive, and comprehensive. We turned these dimensions into statements and presented as 10-point scale (1-low quality, 10-high quality) questions for external raters. We also asked the raters to score the overall discussion quality from each interface condition. For a consistent evaluation environment, all interface enhancements were removed to only display the comments.

Results

Overall, participants reported limited interest in and knowledge of topics covered in articles. For 7-point scale questions, the mean self-reported interest levels were 4.03 (SD=1.48), 3.92 (SD=1.74), and 3.13 (SD=1.61), and knowledge levels were 2.47 (SD=1.52), 2.62 (SD=1.76), and 2.28 (SD=1.19) for baseline, fact-checking, and Factful groups, respectively.

We now present a summary of participants' discussion activities from the study. Table 2 shows per person activity counts. While the total number of comments is not significantly different across interface conditions, fact-checking requests and fact-checking were used more frequently in the Factful group than the fact-checking group. Note that fact-checking requests are a lighter form of contribution than comments, and fact-checking requires much more effort.

activities/person	**Baseline**	**Fact-checking**	**Factful**
Comments	10.5	10.85	10.54
Fact-checks	-	0.58	0.83
Fact-check requests	-	2.75	3.25
Comment "likes"	5.67	6.83	4.58
Total activities	16.17	21.92	20.08
Thread length in words	13.04	10.18	14.37

Table 2. Summary of discussion-related activities during the lab study

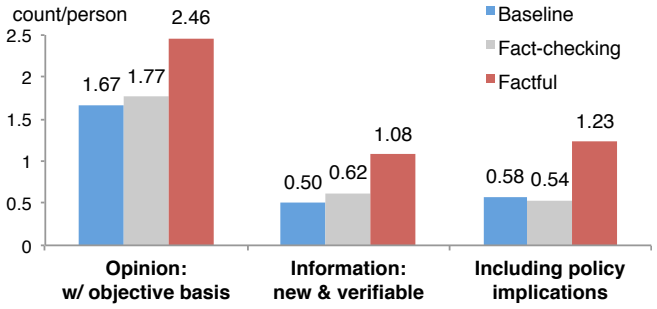

Figure 7. Summary of discourse analysis, showing per-interface comparisons for certain comment types.

The amount of activity is probably not informative enough because people were in a lab study with time limits and external incentives, not completely voluntarily participating. We instead focus on the differences between activity types and discussion quality, which we expect will depict the differences between the conditions more accurately.

Results of our discourse analysis are presented in Figure 7. Of the three labels presented, opinions with objective basis and new and verifiable information are derived from Table 1. We additionally counted comments that included policy implications. Factful users made more comments in all three labels presented, although the differences were not statistically significant (Kruskal-Wallis Test, $p > 0.2$).

The distribution of comment types in each condition is presented in Table 3. The Factful group created more questions and less simple statements than the other groups.

H1 (Readers will hold a more critical view on the article with Factful.) We asked a 7-point Likert-scale question (1-not at all, 7-very much) on how much they trust the article after each session. Average score was lower in Factful (4.03) than in baseline (5.08) and fact-checking (4.77) (Mann-Whitney Test, $p < 0.01$, $z = 2.19$ for baseline and $p < 0.05$, $z = 3.29$ for Fact-checking). Also, external raters' discussion quality score for the Factful group was higher (6.6) than baseline (4.93) and fact-checking (5.67) groups ($p < 0.05$, χ^2=6.97).

While the number of critical opinions per person was not significantly different between three groups (5.14, 4.15, 5.54 per person), the subjects of critic were different between groups. While Factful participants left 10 (0.77 per person) critics about articles, no baseline participant cast doubt on the ar-

count/person	Baseline	Fact-checking	Factful
Opinion	6.25	5.23	6.23
Information	0.83	1.00	1.08
Question	2.00	2.23	3.08
Simple statement	1.42	1.85	0.69

Table 3. The number of comments per each participant, for each comment type and interface condition.

ticles. Instead of criticizing the articles, baseline participants criticized each other (1.25 per person) more than the other two groups did (0.31 for fact-checking and 0.46 for Factful).

H2 (Readers will discuss with more fact-based statements with Factful.) Factful participants discussed with more objective supporting arguments (2.46 per person) than the other two groups (1.67 for baseline, 1.77 for fact-checking). Additionally, they raised more questions asking for objective information (2.54 questions per person) than the other two groups (1.75 for baseline, 1.77 for fact-checking). In addition to simply reading or asking for facts, Factful participants introduced out-of-article information by extending what was available in the article. Although they wrote information-providing comments (1.08 per person) only slightly more than the two other groups (0.83 for baseline and 1.00 for fact-checking), almost half of those were objective information from external sources. Comments from the other two groups were more focused on existing information and unverifiable information.

H3 (Readers will build a rich context surrounding the relevant budget facts using Factful.) The Factful group engaged in discussions with more policy implications than the other two groups. More than just criticizing the government policies and budget spendings, Factful participants came up with new ideas that would be helpful for policy makers. They suggested more opinions with policy implications (1.23 per person) than the other two groups (0.58 for baseline, 0.54 for fact-checking), although the difference was not statistically significant. External raters gave a significantly higher score (6.93) to discussions in Factful for the criterion 'they discussed with various perspectives and supporting grounds' (Kruskal-Wallis Test, $p < 0.05$, χ^2=7.71).

H4 (Readers will initiate and engage in the discussion with more existing discussion activities.) For all participants in all groups, the self-reported interest level increased more after the third session (4.76) than after the first (4.07), although the difference was only marginally statistically significant (Mann-Whitney Test, p=0.06, z=-1.91). If the trend continues with more users, the exposure to other users' accumulated activities may affect the commenting behavior. Being able to initiate discussions not only from the article but also from other users' comments or fact-checking requests can also lead to more active discussions.

DISCUSSION

The role of contextual information: Eight of 13 participants using Factful said that the similar budget-sized program information from automated annotations served as a basis of evaluating government spending. Factful users commented that *"Without such information, it would be hard to determine if the given government spending is worth or not."* [Factful group, p07], and *"By seeing that information, I was able to critically evaluate the government spending with relevant budgetary issues, not blindly agreeing with the article."* [Factful group, p02] Six participants said that the sectoral budget information gave them a basic idea of the article topic and related budgets. One participant noted that the sectoral budget information in Factful made reading through the article easier, because he felt the budget terms and numbers in the article were less obscure.

Expertise in fact-checking: Our study results suggest that participants with low awareness still participated in fact-checking (0.83 times per article per person with Factful). While some fact-checking activities would require domain expertise and reviewing multiple sources, we observed many simple activities such as adding missing references and verifying factual information from a single source. Also, readers not knowledgeable about budgets may still be domain experts related to certain budget programs and do fact-checking.

Generalization to other datasets and countries: In this paper we used news articles and budget structure of South Korea, and a question remains – how the findings might generalize to different cultures. As previous research shows that we can identify common generic budget processes regardless of the legal structure [11], we believe many countries share similar issues in their budget and related processes.

Limitations

Our study mainly focused on understanding the feasibility of our ideas in the context of Factful. The lab study had several limitations. First, study participants were mostly students at a university. Further study is needed to see if the study findings can generalize to taxpayers of varying political standpoints and socio-economic status. Also, the lab setting may have biased the types of social interactions that took place. As more taxpayers engage in open online discussions, quality control and information overload can be issues. When applied to opinionated claims, social voting and the wisdom of the crowd may have a limited effect for quality control [13]. We will consider novel quality control mechanisms and ways to simplify the view.

CONCLUSION AND FUTURE WORK

Despite the importance of a government budget for taxpayers, they have low interest and awareness in budgetary issues and discussions. This paper explored the idea of embedding the contextual budget information and fact-checking layer into news article reading, with the goal of engaging taxpayers in the budgetary discussions online. We implemented this idea with Factful, a web-based newsreading interface powered by the budgetary facts of the Seoul metropolitan government.

For future work, we plan to deploy Factful as a live system and study how taxpayers' budgetary awareness and discussion behaviors change. We will also consider a publisher's and journalist's perspective, as Factful may be useful in data journalism and internal fact-checking for newsrooms. For

generalization, we will explore other civic domains that are rich in factual, structured data such as budgeting process and auditing, as well as healthcare (e.g., disease and nutritional facts). With scalable UIs and techniques to present budgetary facts to more taxpayers, Factful presents a simple yet powerful model for leveraging open government data to support a deliberative democracy online.

ACKNOWLEDGMENTS

This research is supported by the Undergraduate Research Program and High-Risk High-Return Program at KAIST. Juho Kim is partly supported by the Samsung Fellowship.

REFERENCES

1. Brush, A. J. B., Bargeron, D., Grudin, J., Borning, A., and Gupta, A. Supporting interaction outside of class: Anchored discussions vs. discussion boards. In *CSCL'02* (2002), 425–434.
2. Caumont, A. 12 trends shaping digital news, 2013. Pew Research Center. [Online; posted 16-October-2013].
3. Chen, Y., Yang, J., and Ribarsky, W. Toward effective insight management in visual analytics systems. In *IEEE PACIFICVIS '09* (2009), 49–56.
4. Diakopoulos, N., Goldenberg, S., and Essa, I. Videolyzer: Quality analysis of online informational video for bloggers and journalists. In *CHI '09*, ACM (2009), 799–808.
5. Diakopoulos, N., and Naaman, M. Towards quality discourse in online news comments. In *CSCW'11*, ACM (2011), 133–142.
6. Elster, J. *Deliberative democracy*, vol. 1. Cambridge University Press, 1998.
7. Ennals, R., Trushkowsky, B., and Agosta, J. M. Highlighting disputed claims on the web. In *WWW'10*, ACM (2010), 341–350.
8. Fishkin, J. S., and Luskin, R. C. Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica 40*, 3 (2005), 284–298.
9. Frey, D. Recent research on selective exposure to information. *Advances in experimental social psychology 19*, 1 (1986), 41–80.
10. Garrett, R. K., and Weeks, B. E. The promise and peril of real-time corrections to political misperceptions. In *CSCW '13*, ACM (2013), 1047–1058.
11. Gomez, P., Friedman, J., and Shapiro, I. Opening budgets to public understanding and debate: Results from 36 countries. *OECD Journal on Budgeting 5*, 1 (2005).
12. Graham, J. The reader's helper: A personalized document reading environment. In *CHI'99*, ACM (1999), 481–488.
13. Graves, L. *Deciding What's True: Fact-Checking Journalism and the New Ecology of News*. PhD thesis, Columbia University, 2013.
14. Hullman, J., Diakopoulos, N., and Adar, E. Contextifier: Automatic generation of annotated stock visualizations. In *CHI'13*, ACM (2013), 2707–2716.
15. Iacobelli, F., Birnbaum, L., and Hammond, K. J. Tell me more, not just "more of the same". In *IUI'10*, ACM (2010), 81–90.
16. Kong, N., Hearst, M. A., and Agrawala, M. Extracting references between text and charts via crowdsourcing. In *CHI'14*, ACM (2014), 31–40.
17. Kriplean, T., Bonnar, C., Borning, A., Kinney, B., and Gill, B. Integrating on-demand fact-checking with public dialogue. In *CSCW'14*, ACM (2014), 1188–1199.
18. Kriplean, T., Morgan, J., Freelon, D., Borning, A., and Bennett, L. Supporting reflective public thought with considerit. In *CSCW '12*, ACM (2012), 265–274.
19. Kriplean, T., Toomim, M., Morgan, J., Borning, A., and Ko, A. Is this what you meant?: Promoting listening on the web with reflect. In *CHI '12*, ACM (2012), 1559–1568.
20. Lampe, C., and Resnick, P. Slash(dot) and burn: Distributed moderation in a large online conversation space. In *CHI'04*, ACM (2004), 543–550.
21. Liao, Q. V., and Fu, W.-T. Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions. In *CHI'14*, ACM (2014), 2745–2754.
22. Liaw, R., Zilnik, A., Baldwin, M., and Butler, S. Maater: Crowdsourcing to improve online journalism. In *CHI EA'13*, ACM (2013), 2549–2554.
23. Meola, M. Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *portal: Libraries and the Academy 4*, 3 (2004), 331–344.
24. Metzger, M. J. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology 58*, 13 (2007), 2078–2091.
25. Munson, S. A., Lee, S. Y., and Resnick, P. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM* (2013).
26. Otterbacher, J. 'helpfulness' in online communities: A measure of message quality. In *CHI'09*, ACM (2009), 955–964.
27. Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. Understanding the participatory news consumer. *Pew Internet and American Life Project 1* (2010), 19–21.
28. Salton, G., and McGill, M. J. *Introduction to modern information retrieval*. McGraw-Hill, 1983.
29. Wu, Y., Agarwal, P. K., Li, C., Yang, J., and Yu, C. Toward computational fact-checking. *Proc. of the VLDB Endowment 7*, 7 (2014).
30. Zaller, J. *The nature and origins of mass opinion*. Cambridge university press, 1992.