Dike: Crowdsourced Comprehensible Judgment Generation

Hyeungshik Jung
KAIST
hyeungshik.jung@kaist.ac.kr

Dongkwan Kim
KAIST
dong112d@kaist.ac.kr

Hyungyu Shin
KAIST
hyungyu.sh@kaist.ac.kr

ABSTRACT

Reading written judgments requires high comprehension because of complex structure of statements and legal jargon. We introduce *Dike*, a crowdsourcing platform to translate written judgments into more comprehensive reading. We propose 4 stage crowdsourcing workflow: split - polish - connect - revise. A statement in written judgment is split in the split phase, and each split part becomes a complete sentence in the polish phase. After connecting the complete sentences by adding appropriate connectors to each sentence, sentences are revised for better readability. We deployed our system and 17 users have used our platform. Results show that Dike translates original written judgment into more comprehensive reading.

INTRODUCTION

The written judgment is the most important and basic legal text about interpreting the law based on the events between various stakeholders. Nonetheless, the readability of judgments is lower than other kinds of texts [3]. This is not only important for litigants to understand their own interests, but also for the public right to access legal information, prevention of distorted or exaggerated press reports about judgment, above all, the trust to the judiciary as a constituent of democracy [1, 6].

Written judgment often contains complex statement structures and jargon because the meaning should not be misinterpreted, which makes it hard to understand for the reader. Statistics [1] in South Korea show that 67.6% of readers need 2 or 3 times of reading to understand written judgments. It indicates that people are able to understand the written judgment, but it is not a straightforward task. We envision that crowd participates in the procedure of translating the written judgment into better reading.

Figure 1 represents the process for the translation of a statement a in written judgment. Our crowdsourcing workflow consists of 4 phase: split - polish - connect - revise workflow. A worker is allowed to go through the whole phase sequentially. For a statement that is hard to understand, the worker first split the statement into multiple pieces. It makes worker to easily concentrate each meaningful unit in the statement. Next, the worker makes each split into a complete statement in the polish phase. The complete sentences are connected by appropriate connectors in the connect phase. Finally, worker revises each sentence into a more readable way.

We allow multiple workers to contribute in a single phase for better quality. Similar to a version control system, each contribution of worker make a branch and compete to get the best result. Figure 3 shows the tree structure of overall phase. To reduce the size of tree, we adopted natural selection algorithm for cutting branches with low quality.

We deployed our system and 17 users have used our platform. The results show that our crowdsourcing workflow successfully translates written judgments into more readable statements.

BACKGROUND AND RELATED WORK

Research and statistics show that written judgment is difficult to understand for a citizen. Research [3] shows that written judgment is difficult to understand for citizen considering various criteria. Survey results [2] of The Law Times show that 70% of 130 respondents pointed out the use of too long sentences as a reason for difficulty. Also, in the same survey, 20% of respondents said that use of pedantic terminology is also another reason. In our teammate's experience at The KAIST Times, he reported that among many kinds of texts he had to read to write articles, it was especially hard to write articles about written judgments. He pointed out that the continuation of sentences that are hard to understand at one time gives a high reading cost to the understanding of the whole text.

SYSTEM

Interface

Figure 1 shows our interface *Dike*. It supports three different tasks: A user can 1) read more comprehensive sentences for difficult sentences, 2) see how the difficult sentences are translated, and 3) browse which part of a written judgment is hard to understand. The target user of task 1) is a normal citizen, that of 2) is a judge who wants to know why citizens have a difficulty on understanding, and that of 3) is people who want to contribute in translation but hard to know which part of written judgment needs contribution. To support the tasks, Dike provides perceived difficulty for each part of a written judgment and translated text with more comprehensive information. In the next section, we describe the crowdsourcing workflow to support the tasks.

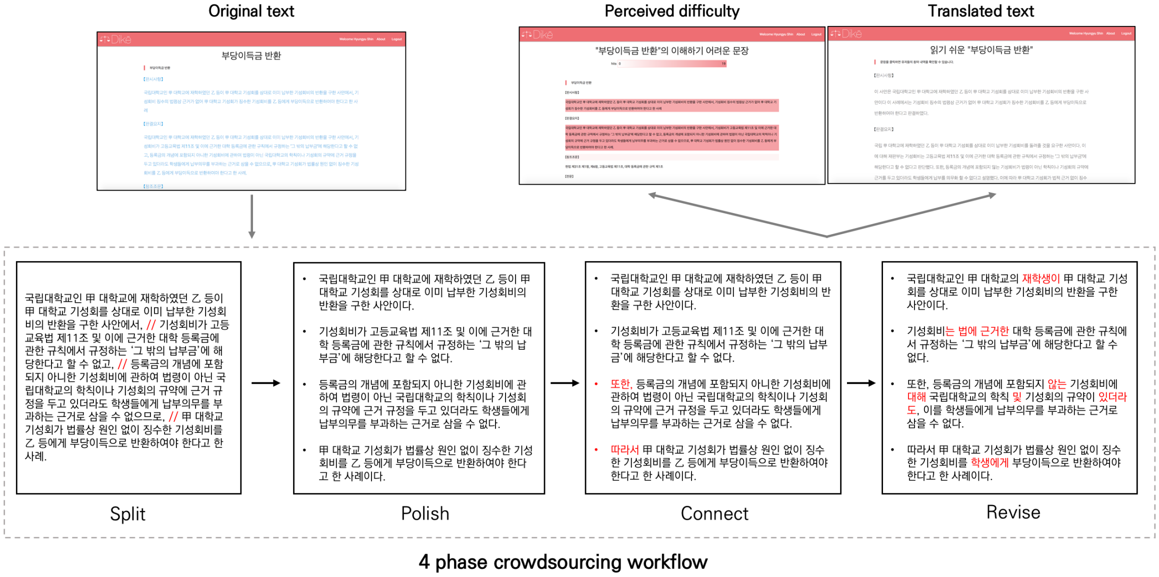
Workflow

Figure 1 represents our crowdsourcing workflow, which consists of 4 phases: split - polish - connect - revise. The key idea is the divide and conquer approach. In other words, crowd split a sentence that is difficult to understand and translate each piece into more readable sentences.

As we have 4 phases, which is a relatively heavy task that could cause people to stop performing the tasks, we need a

---

[1] http://www.hankookilbo.com/v_print.aspx?id=ede600cfe8264610b5eea419d8fe56dc


Original text
Perceived difficulty
Translated text
부당이득금 반환
"부당이득금 반환"의 이해하기 어려운 문장
읽기 쉬운 "부당이득금 반환"
Split
Polish
Connect
Revise
4 phase crowdsourcing workflow

국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안에서, // 기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없고, // 등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 없으므로, // 甲 대학교 기성회가 법률상 원인 없이 징수한 기성회비를 乙 등에게 부당이득으로 반환하여야 한다고 한 사례.

- 국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안이다.
- 기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없다.
- 등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 없다.
- 甲 대학교 기성회가 법률상 원인 없이 징수한 기성회비를 乙 등에게 부당이득으로 반환하여야 한다고 한 사례이다.

- 국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안이다.
- 기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없다.
- 또한, 등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 없다.
- 따라서 甲 대학교 기성회가 법률상 원인 없이 징수한 기성회비를 乙 등에게 부당이득으로 반환하여야 한다고 한 사례이다.

- 국립대학교인 甲 대학교의 재학생이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안이다.
- 기성회비는 법에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없다.
- 또한, 등록금의 개념에 포함되지 않는 기성회비에 대해 국립대학교의 학칙 및 기성회의 규약이 있더라도, 이를 학생들에게 납부의무를 부과하는 근거로 삼을 수 없다.
- 따라서 甲 대학교 기성회가 법률상 원인 없이 징수한 기성회비를 학생에게 부당이득으로 반환하여야 한다고 한 사례이다.

Figure 1: Overview of the system

quality control mechanism. we permit users not to follow all the phases. Users can stop proceeding the task whenever they want. Then another user can pick the intermediate result and finish it. In summary, we maintain a tree structure for users work like a version control system. The system will give a starting point to a user, and the user simply follows the phases from the point. In the following subsections, we will describe our crowdsourcing workflow and quality control.

Split - Polish - connect - Revise workflow

Split

Crowd first split a sentence into multiple pieces. As we split the sentence, a reader can concentrate on meaningful pieces, which enhances the readability. Figure2 shows the interface for splitting. A user can specify boundaries between split via clicks.

Polish

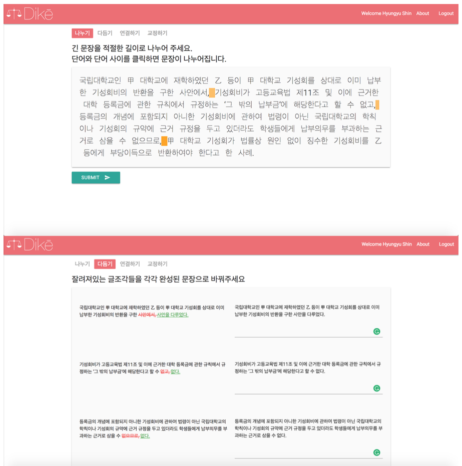
Each split is not a complete sentence because crowd simply split the sentence. To make a complete sentence, crowd polish each split into a complete sentence. Figure 2 shows the interface for polishing. As user modify the sentences, we show the difference between the original sentence and polished sentence to show how the user modify the sentence.

Connect

With complete sentences, the next task is to connect the sentences by adding appropriate connectors. It makes sentences more naturally readable, which further enhances readability. The interface of connect phase is the same as that of polish phase.

Revise

Finally, crowd can revise each complete sentences into more easy sentences. Written judgments normally contain multiple jargons, which hinders readability. Revising the sentences such as substituting jargon into more easy words or paraphrasing the whole sentence also enhance readability. The interface of revise phase is also same as that of polish phase.



긴 문장을 적절한 길이로 나누어 주세요.
단어와 단어 사이를 클릭하면 문장이 나누어집니다.

국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안에서, 기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없고, 등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 없으므로, 甲 대학교 기성회가 법률상 원인 없이 징수한 기성회비를 乙 등에게 부당이득으로 반환하여야 한다고 한 사례.

잘려져있는 글조각들을 각각 완성된 문장으로 바꿔주세요

국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 ~~사안에서,~~ 사안을 다루었다.	국립대학교인 甲 대학교에 재학하였던 乙 등이 甲 대학교 기성회를 상대로 이미 납부한 기성회비의 반환을 구한 사안을 다루었다.
기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 ~~없고,~~ 없다.	기성회비가 고등교육법 제11조 및 이에 근거한 대학 등록금에 관한 규칙에서 규정하는 '그 밖의 납부금'에 해당한다고 할 수 없다.
등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 ~~없으므로,~~ 없다.	등록금의 개념에 포함되지 아니한 기성회비에 관하여 법령이 아닌 국립대학교의 학칙이나 기성회의 규약에 근거 규정을 두고 있더라도 학생들에게 납부의무를 부과하는 근거로 삼을 수 없다.

Figure 2: The split interface (top). A user can click boundaries between split. The polish interface (bottom). It shows how the user modify each sentence.

We get a tree structure as we go through the 4 phase workflow. Figure 3 represents the tree structure. In the following subsection, we describe why we need quality control mechanism and how to achieve it.

Quality control

Drawbacks of the workflow

To get one complete result, dividing tasks and distributing them to the crowd reduces the load on the crowd, but we found

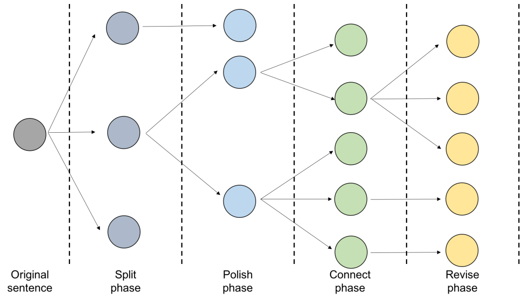


Figure 3: The tree structure from our workflow.

from the user study that there are some side effects. First, there are many steps to take, so it takes a long time to get one result. Second, the degree to which quality control is required varies depending on the requirement of each step. Third, as the crowd does the task at intervals, the general voting method does not give a fair evaluation to candidates made earlier and candidates made later.

We suggest applying population model of growth [5] to each intermediate result of our workflow.

Solution: population model of intermediate results

In population model of growth, dP/dt, differential equation for the population $P(t)$ at time t is:

$$\frac{dP}{dt} = r \cdot P \cdot \left(1 - \frac{P}{K}\right) \tag{1}$$

Maximum growth rate r and carrying capacity K are constant in this model. This model means that the growth rate increases initially, decreases after the population reaches half of K which is the limit nature can accommodate, eventually the population converges to K. In the wild nature, natural selection occurs depending on the value of r and K given to each species [4].

Our solution starts by describing the carrying capacity k as a function of the number of votes v that a candidate received and the total number of votes v_t. This equation is designed to give k, which are more favorable to population growth, as more candidates receive more votes.

$$k(v, v_t) = K \cdot \frac{v}{v_t} \tag{2}$$

In addition to modifying K to $k(v, v_t)$, we also set r as a function of v_t. Since the result of our workflow changes by vote continuously, unlike r in the real world which is determined as a constant, the initial r may not reflect the quality of the actual candidate. Therefore, to reduce the initial effect, $r(v_t)$ becomes the product of the constant R_{max} and $v_t/V_{slowstart}$ before v_t reaches $V_{slowstart}$. If $v_t \geq V_{slowstart}$, v_t will be R_{max}. This term can moderate the misjudged natural selection in the beginning.

$$r(v_t) = R_{max} \cdot min\left(\frac{v_t}{V_{slowstart}}, 1\right) \tag{3}$$

Now, we can get the population $n(t)$ calculated as the result of $r(v_t)$ and $k(v, v_t)$. In our model, $n(t)$ will be the quality metric of the candidates at the same stage, not just the number of received votes.

$$\frac{dn}{dt} = r(v_t) \cdot n \cdot \left(1 - \frac{n}{k(v, v_t)}\right) \tag{4}$$

Real application

There are some constants that need to be determined in advance to apply this model. They are not only K, R_{max}, $V_{slowstart}$ in equation 2 and 3, but also N_{init}, the initial value of n and $N_{extinct}$, the minimum requirement of the population to be the candidate. That is, every candidate starts with population N_{init}, will be removed when their population is less than $N_{extinct}$.

We determined these constants heuristically. Assuming that K and R are 10 and 2 respectively, we performed a simulation which the first candidate wins with 2/3 probability in pairwise voting. As a result of this simulation, we can conclude that the combination of $V_{slowstart} = 10$, $N_{init} = 2$, $N_{extinct} = 1$ makes converged result of final workflow within 20 virtual users.

$$K = 10,\ R_{max} = 2,\ V_{slowstart} = 10,\ N_{init} = 2,\ N_{extinct} = 1 \tag{5}$$

Also, we limited the maximum number of candidate pool to five. At this time, the pool includes only candidates that have populations more than $N_{extinct}$.

After a total of five candidates registered and only one candidate has more populations than $N_{extinct}$ by natural selection, our system deems the candidate the best for that step.

Expected effect

Since the initial growth of the population is slower than 'linear', even the candidate generated later can be chosen as the best candidate. Also, because the population reduction of candidates with low quality is faster than linear, the system can remove candidates with low quality efficiently.

EVALUATION

Our pilot study sought to answer whether Dike can reduce the complexity of judgment document, as well as understand how our natural selection algorithm works in the real world.

Procedure

We chose a Korean judgment on school support fees [2] as our target judgment. Students attending national universities sued their universities for charging school support fees in addition to tuition fees, and the court sentenced that additional fee is illegal. We chose this judgment because our participants are likely to be interested.

We recruited total 17 users among researchers' acquaintances and students from the school bulletin board. They were given a short instruction about the project and interface and asked to work on any sentences they felt difficult to understand. As a result, our participants did total 246 tasks including split, polish, connect, revise steps, and did 398 times of voting.

Results: Change in sentences

We analyzed how our system changed the original judgment.

We compared the distribution of word count in each sentence of original judgment and judgment processed by our workers

(Figure 4). We verified the average word count in sentence decreased from 130 to 75, and the longest sentence in processed judgment contains 221 words while the counterpart of original judgment contains 540 words.

We also checked whether processed sentences preserve the meaning of sentence in original judgments. While we did not run an in depth-comparison between sentences, we found there was not any significant change in the meaning of sentences.

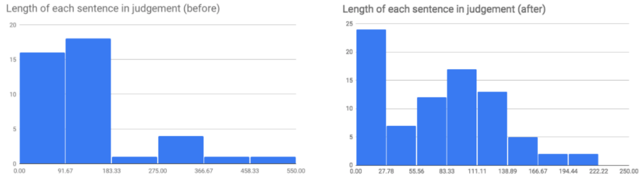


Figure 4: Distribution of word counts in original judgment (left), and processed judgment by the system (right)

Results: Effect of natural selection

The rationale behind the design of natural selection algorithm was to ensure that the result of later revisions are not dominated by earlier revisions, and can be selected if they get a lot of votes. We could found the example cases where the later revision get selected against earlier revisions. In Figure 5, the revision number 3423 (orange line) beat two earlier revisions as it gets a lot of votes. In Figure 6, a revision from malicious user (orange line). was selected as best revision (at phase 2), however, it was soon replaced by other revision (blue line).

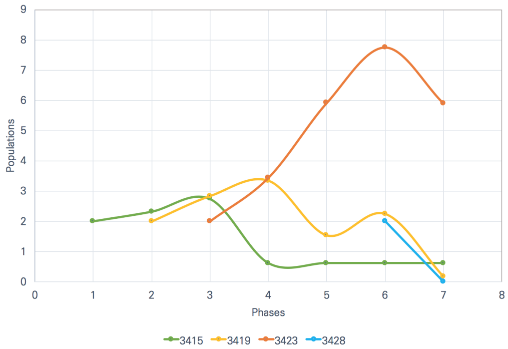


Figure 5: Change of computed populations for various revisions for a single sentence.

DISCUSSION

While the result of our pilot study reveals the possibility of crowdsourced, easy-to-read judgment documents, our preliminary result reveals several discussion points as well.

Lowering the bar of the task

While we split the revision task into 4 steps, participants in pilot study still found the task difficult. The most frequent difficulty was jargons used in judgment. Providing adequate assistance for jargon will ease revision task.

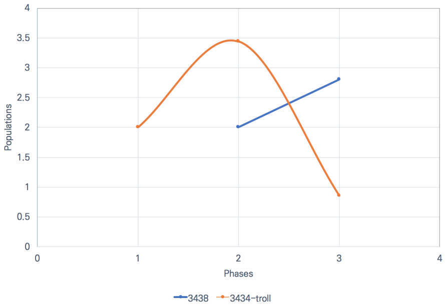


Figure 6: Revision by malicious user was replaced by revision with further worker by voting mechanism

Designing more robust natural selection algorithm

Current natural selection algorithms depend on three parameters, and we had to fine-tune parameters that make convergence within 20 participants. Improving the algorithm to calculate parameters analytically according to the number of required participants will increase the applicability of the proposed natural selection algorithm.

Exploring diverse representation of judgment document

In this work, we provided the only linear representation of judgment sentences for participants. However, our participants frequently expressed needs for alternative representations, such as representation for parallel structure and nested structure.

REFERENCES

1. Michael Curtotti, Wayne Weibel, Eric McCreath, Nicolas Ceynowa, Sara Frug, and Tom Bruce. 2015. Citizen science for citizen access to law. *J. Open Access L.* 3 (2015), 57.
2. Seoul Central District Law. 2014. Return of Unfair Benefit. 2014GaHap3141 (2014). `http://www.law.go.kr/LSW/precInfoP.do?precSeq=174936&mode=0`.
3. Joo Yong Park Min Jo Ko. 2015. A Research on the readability of sentencing. *THE KOREAN JOURNAL OF FORENSIC PSYCHOLOGY* 6 (2015), 33–52.
4. Eric R Pianka. 1970. On r-and K-selection. *The American Naturalist* 104, 940 (1970), 592–597.
5. Pierre-François Verhulst. 1838. Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a. *Quetelet* 10 (1838), 113–121.
6. Yang. 2015. Research on the measures to promote public awareness and understanding regarding the judgments of the courts. (2015).