Joint Transportation Research Program

Principal Investigator: Hubo Cai Purdue University, hubocai@purdue.edu, (765) 494-5028

Program Office: jtrp@purdue.edu, 765.494.6508, www.purdue.edu/jtrp

Sponsor: Indiana Department of Transportation, 765.463.1521

SPR-4422 2020

Automating the Generation of Construction Checklists

Introduction

Construction inspection is a critical component of the quality assurance/quality control (QA/QC) program at the Indiana Department of Transportation (INDOT). In the current practice, inspectors manually gather the construction requirements for a construction activity from the plans and specifications, which is time-consuming and error-prone, and heavily relies on the experience of the inspectors and their subjective interpretation of the textual requirements. Automating the requirements-gathering process and presenting it in a consistent and easy-to-use format can greatly enhance inspection efficiency and accuracy and, consequently, address the resource shortage at INDOT.

The objective of this project is to develop a tool to automate the generation of a customized and dynamic checklist that contains only the applicable, specific information for a construction activity at the pay item level with adequate details to guide the construction inspection process. Besides basic information, the resulting checklist (composed of a list of check items) includes the acceptance criteria, the inspection frequency, and the associated risk. The checklist also has mechanisms that connect individual checklist items to relevant, available training materials that come in varying formats, such as photos, videos, textual documents, and websites. The project scope includes Divisions 200, 300, 400, 500, and 700.

A relational database was designed and constructed in this project to store the pay items and check items as well as the construction process and stages. A graphical user interface and specific tools were developed to perform the following functions: (1) automatic generation of a customized checklist at the pay item level; (2) access to a checklist display aligned with the repetitive/cyclical nature of construction workflows; (3) navigation between cross-referenced check items; (4) subgrouping options based on the responsibility, risk level, and inspection frequency; and (5) linking options to accommodate photos, videos, textual documents, and websites. Natural language processing (NLP) and machine learning algorithms were explored to extract the inspection requirements from textual documents and restructure them into the checklist questions.

Findings

The main findings of this project are as follows:

- The database approach (including both the database design and the user interface) is capable of achieving the desired functionalities to automatically generate customized checklists at the pay item level.
- The division-section-subsection-requirement sentences hierarchy can help organize the construction requirements and the applicable pay items. In particular, subsections play a central role in linking pay items, construction processes/stages, and check items.
- Testing the four mechanisms that could accommodate the inspection documents and training materials in varying formats revealed the following:
  - A URL link is sufficient to connect checklist items with websites.
  - Embedding files in the database is impracticable due to size constraints and data duplication.
  - Organizing files under folders and linking pay items to folders are efficient approaches but they require additional effort to explore and navigate through the file system.
  - The variant version, wherein each check item is linked to a central linking file that contains links to all the applicable documents stored on the cloud, will save storage space and eliminate data redundancy. This mechanism is recommended to INDOT for implementation.

- NLP techniques developed for the automatic extraction of the check items from the INDOT Standard Specification 2018 achieves an approximate accuracy of 90%.
- A regular expression (RE) is efficient in identifying the hierarchy (subsections-section-division) of a file and separating sentences based on their textual patterns.
- The sentence classifier composed of GloVe (a word embedding algorithm) and a convolutional neural network (CNN) (a deep learning algorithm) results in the highest accuracy and the least loss.
- Using training samples from all divisions, a higher percentage of data, and/or a validation set improves the classification performance.
- A valuable dataset that is comprised of 3,171 sentences extracted from the INDOT Standard Specification 2018 and labeled with the correct classes ("1" for requirement and "0" for nonrequirement) can be used to train models and automatically extract inspection requirements from other versions of INDOT's specifications to address the specification modifications issue.

Implementation

The main deliverables of this project are the relational inspection database and NLP algorithm. There are at least three options for implementing the database.

1. The database can be directly used as a stand-alone system. A user can import the pay items from the INDOT contract information book (CIB) into the database and use the already-developed user interface to generate a customized checklist for the chosen pay item(s) and exporting it in varying formats (Excel, Word, PDF).
2. The database can be connected to the field inspection and documentation tool being developed. When an inspector chooses a specific pay item, the applicable construction requirements can be extracted from the database. The inspection results are then saved and stored together with the quantity check.
3. The database can be imported into the enterprise database at INDOT and used either as a standalone application or connected to the field inspection and construction documentation system.

The NLP algorithm can achieve about 90% accuracy. The classification model was trained based on INDOT's Standard Specification 2018 but can be directly used on other versions to extract construction requirements automatically. Users are cautioned that the scope of this project included Divisions 200, 300, 400, 500, and 700, and thus only applies to pay items and construction requirements in those divisions.

At the time of this writing, a team composed of experts from the Management Information System (MIS) unit and the Construction group at INDOT is working on the system implementation.

Recommended Citation for Report

Cai, H., Jeon, J., Xu, X., Zhang, Y., & Yang, L. (2020). *Automating the generation of construction checklists* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2020/23). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317273

View the full text of this technical report here: https://doi.org/10.5703/1288284317273

Published reports of the Joint Transportation Research Program are available at http://docs.lib.purdue.edu/jtrp/.