TriDurLE

National Center for Transportation Infrastructure Durability & Life-Extension

Project ID: 2021-WSU-WVU-01

DEVELOPMENT OF INFRASTRUCTURE RESEARCH WEEKLY APP

Final Report

By

Kakan Dey, Ph.D., P.E. (ORCID ID: 0000-0002-5875-6180)
Assistant Professor, Department of Civil and Environmental Engineering
West Virginia University, Morgantown, WV 26506
Phone: 304-293-9952; Email: kakan.dey@mail.wvu.edu

Xianming Shi, Ph.D., P.E. (ORCID ID: 0000-0003-3576-8952)
Professor and Chair, Department of Civil and Environmental Engineering
Washington State University, Pullman, WA 99164, USA.
Phone: 1-509-335-7088; Email: xianming.shi@wsu.edu

Md Tanvir Ashraf
Graduate Research Assistant
West Virginia University

Ajay Inturi
Graduate Research Assistant
West Virginia University

for

National University Transportation Center TriDurLE
Department of Civil & Environmental Engineering
405 Spokane Street, PO Box 642910
Washington State University, Pullman, WA 99164-2910

April 2023

Acknowledgments

This project was funded by the National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE @ WSU), one of the National University Transportation Centers funded by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are also very grateful to DOT personnel, researchers, and industry professionals for their participation in the study.

Disclaimer

Table of Contents

List of Figures

List of Tables

Executive Summary

The poor condition of America's transportation infrastructure calls for immediate investment, cost-effective innovations, and a holistic solution to enhance its durability and life-extension. Many transportation agencies and research organizations, including TriDurLE, have been conducting innovative research to improve the durability and life-extension of transportation infrastructure. However, research findings from articles/reports originated from numerous studies by diverse organizations, and relevant professional activities (e.g., seminars, webinars, and workforce development opportunities) were not aggregated at a single platform for convenient access and broader dissemination. Due to the proliferation of smartphone devices, an App that compiles the advancements in research and practices related to durability and life-extension of transportation infrastructure and other professional development opportunities will be useful to diverse, relevant stakeholders. The goal of this study was to develop a smartphone-based "Infrastructure Research Weekly" App that shares the curated most recent research study findings (e.g., reports, journal/conference articles), news on transportation infrastructure durability and serve as the "One-Stop-Shop" for researchers from the consortium universities, industry practitioners and professionals and students. The App also presents information on seminars, training, and other career opportunities related to transportation infrastructure durability and life-extension. Six interrelated tasks were executed to achieve the research goal. A stakeholder survey was conducted to assess the needs and expectations of the relevant stakeholders from the App.

The primary content of the App was collected from major journals, magazines, websites, news media, TRID database, and social media via keyword searches, using web scraping or artificial intelligence (AI) techniques. This study employed keyword search to collect literature published between 2020 and 2022 from several bibliographic databases. The collected literature was filtered by applying the PRISMA framework. The Structural Topic Modeling (STM) method was used to find the major research topics based on the relative frequencies of the keywords used in the abstracts of the collected literature. The STM method revealed 12 topics in transportation infrastructure durability and life-extension, which include the application of artificial intelligence (AI) methods, research on innovative materials, performance monitoring, maintenance, repair, and rehabilitation (MR&R) activity selection and optimization, and effects of natural hazards and climate change impacts. Expected topic proportions showed that research on the application of deep learning increased over the last three years. Emerging trends in transportation infrastructure durability and life-extension research considering the growing risks of climate change and extreme weather events promise cost-effective solutions in improving transportation infrastructure durability and life-extension.

Chapter 1. Introduction

1.1 Research Background

Efficient and reliable transportation systems are the backbone of a nation's economy. The length of public roadways in the U.S. is over 4 million miles, with about 614,387 bridges (ASCE Infrastructure Report Card, 2021). Although the U.S. has the largest transportation infrastructure system in the world, the surface transportation system infrastructure is in poor condition. According to the American Society of Civil Engineers (ASCE), the current rating of U.S. roads and bridges are D and C, respectively (ASCE Infrastructure Report Card, 2021). An estimated immediate investment of U.S. $786 billion is needed to improve the condition of the U.S. surface transportation system (ASCE Infrastructure Report Card, 2021). In 2015, Fixing America's Surface Transportation (FAST) act was signed into law to ensure the long-term funding availability for U.S. transportation systems (FAST Act Summary, 2016). To promote the U.S. economic competitiveness and quality of life, the development of advanced, innovative, and cost-effective solutions to improve the condition and durability are critical for the aging transportation infrastructure in the U.S.

Recently, many transportation agencies and research organizations, including National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE), have been conducting cutting-edge research on cost-effective and holistic solutions that enhance multimodal infrastructure durability. The TriDurLE was established in 2019 with a vision of improving the durability and extending the transportation infrastructure life through research, education and diversity, workforce development, technology transfer, and partnerships. TriDurLE funded projects in various sub-fields of infrastructure durability and life-extension research (TriDurLE, 2021). Creating a platform that compiles the most recent research advancements and other opportunities (e.g., seminars, workshops, job opportunities) is necessary to encourage agencies, professionals, and researchers to adopt new/innovative, and cost-effective best practices. The TriDurLE's research activities have been focused on six interrelated thrust areas (https://tridurle.wsu.edu/research/), which are supported by the educational and technology transfer activities to maximize synergies and adaptability across multiple transportation modes and among diverse stakeholders (TriDurLE, 2021). A smartphone App has great value in compiling the most recent advancements in transportation infrastructure durability and life-extension research and practices and sharing a weekly digest with relevant researchers from the TriDurLE consortium universities, industry practitioners and professionals, and students. This project aims to develop an App that compiles relevant news, publications, seminars, training, and other opportunities in infrastructure durability and life-extension weekly. This App shares the research activities and findings of TriDurLE and others with relevant faculty of the consortium universities, industry practitioners and professionals, and students weekly.

1.2 Research Objectives

This project supports the TriDurLE research goals by conducting educational and technology transfer activities to maximize the dissemination of research findings and practices in infrastructure durability and life-extension conducted by the researchers at the TriDurLE and other research organizations around the world. The goal of this project was to develop an App to share the transportation infrastructure durability-relevant research products, findings, practices, education, and professional training opportunities, and other resources available weekly with relevant faculty, industry practitioners/professionals, and students. The specific objectives of this study were to: (1) Conduct an extensive review of the relevant data sources and APP technologies; (2) Survey stakeholders to identify users' needs and expectations; (3) Develop a conceptual framework of the App; (4) Develop an algorithm to filter and categorize the relevant contents effectively; (5) Perform pilot testing to facilitate the deployment; and (6) Develop final report and App.

1.3 Report Overview

The rest of this report is organized as follows: Chapter 2 on "Research Methods" provides a comprehensive overview of the adopted research methods for data collection, research trend analysis, and App development procedure. Chapter 3 explains the result from the literature collection, research trend analysis, and App development. Finally, Chapter 4 presents the concluding remarks.

Chapter 2. Research Methods

This chapter presents the methods adopted for data collection, trend analysis, and App development.

2.1 Data Collection

This study focused on infrastructure durability and life-extension research to identify current practices and state-of-the-art. This study utilized a systematic literature review approach to reduce biases and omissions of the traditional literature review approach to improve quality and transparency. This systematic review was performed by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Liberati et al., 2009). We conducted a literature search to identify research papers, project reports, and book chapters related to infrastructure durability and life-extension. More specifically, all the bibliographic references published in the last three years in ScienceDirect, Springer, and Transportation Research International Documentation (TRID) database were collected for the topic modeling and systematic literature review. The literature search was confined to the last three years as this study focused on the recent trend in research and practices. A brief overview of the steps followed in this study is presented in Figure 1, and the steps are discussed in subsequent sections.

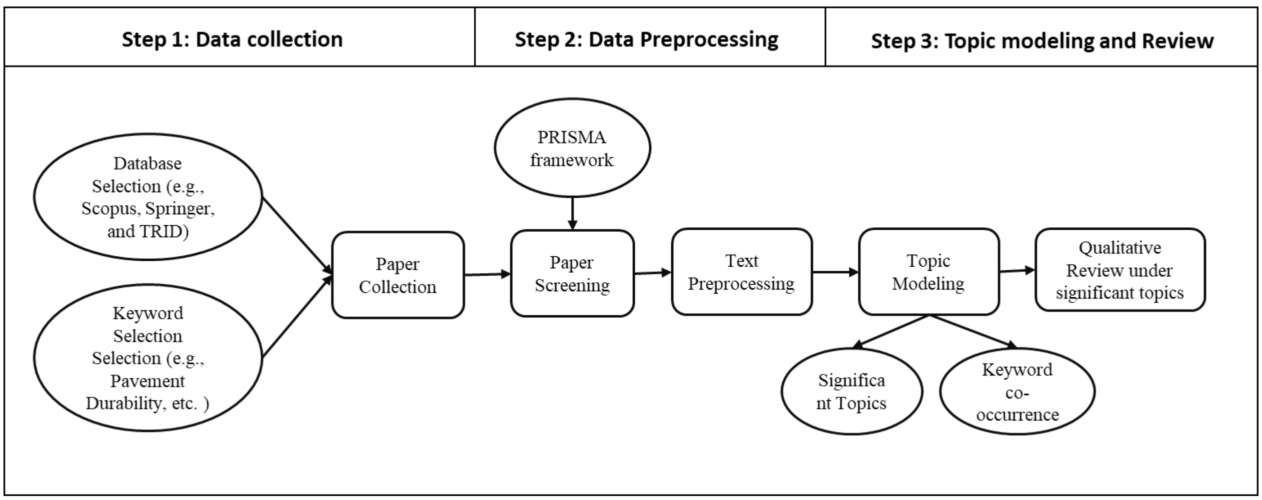

Figure 1: Method for systematic text analysis

2.1.1 Overview of the PRISMA framework

The PRISMA is an established protocol for performing systematic reviews and consists of a 27-item checklist (Liberati et al., 2009). The PRISMA framework was developed by a group of 29 scholars in the medical field with the intent to increase the transparency and accuracy of literature reviews (Pahlevan-Sharif et al., 2019). The flow diagram of the PRISMA framework is shown in Figure 2. At first, studies are selected from the database search using relevant keywords. After collecting the literature, duplicates, and papers not meeting eligibility criteria are removed. Description of the 27-item checklist of the PRISMA framework can be found in Liberati et al. (2009). The PRISMA framework was chosen for this study over other existing protocols because of its comprehensiveness, used in several disciplines, and decrease bias and omissions (Liberati et al., 2009). Some features in the PRISMA framework (e.g., items related to the risk of bias, confidence intervals, measures of consistency, sensitivity or subgroup analyses, outcome level assessment, and combining results of meta-analysis studies) were not used in this study as they were not applicable to the objectives of this study.

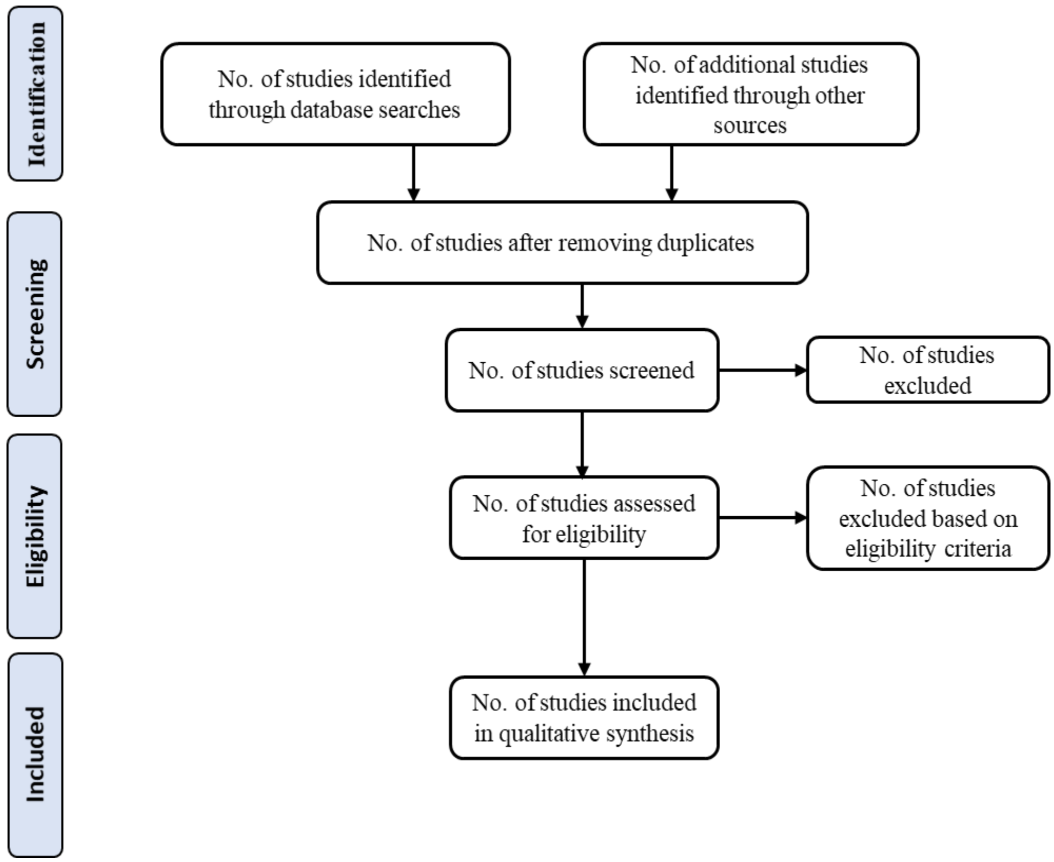

Figure 2: The PRISMA approach flow diagram for qualitative synthesis (Adopted from Liberati et al., 2009).

2.1.2 Literature collection and preprocessing

Three different databases were scraped to capture a wide range of studies related to transportation infrastructure and durability. The bibliographic databases considered in this study were- (1) Scopus (ScienceDirect, Elsevier), (2) SpringerLink (Springer), and (3) TRID database. The selection of keywords was crucial in capturing relevant studies related to the topic and reducing omissions and biases. The keywords were selected based on the subject knowledge to capture documents related to transportation infrastructure durability and life extension. The selected keywords were "(Transportation infrastructure OR bridge OR pavement OR road OR tunnel) AND (service life OR durability OR preservation OR rehabilitation OR repair OR life-extension)". Documents published between January 2020 and June 2022 were considered to keep the studies manageable. The collected studies were then sorted and preprocessed based on the inclusion and exclusion criteria. The set of inclusion and exclusion criteria for an article was as follows: the article was written in English; the document was published in a scientific journal, conference proceedings, book, book chapter, or workshop; duplicate articles from several bibliographic databases were considered only once; journal article was considered only if the impact factor of the journal was greater than 1; and dissertations, ongoing research, guest editorials, poster sessions, and blogs were not considered.

From the selected bibliographic databases, the title, abstract, authors' names and affiliation, journal name, and year of publication data for each study was extracted in a RIS format which was then converted into a CSV format for further analysis. The authors then reviewed the collected data to detect any irrelevant papers or papers with incomplete information. In this process, any paper appearing in multiple bibliographic databases was considered only once in our analysis.

2.2 Trend Analysis

2.2.1 Topic modeling

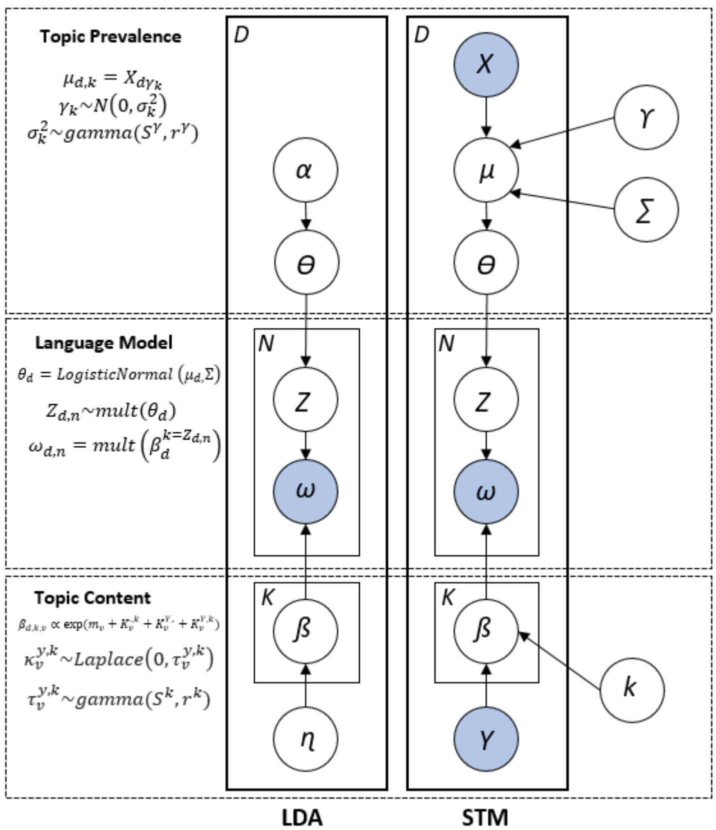

Figure 3: Plate notation of LDA and STM (Adopted from Blei, 2012; Roberts et al., 2016)

Topic modeling is a text-mining technique that examines the "hidden" thematic structure of a set of texts (Bai et al., 2021). Recently, topic modeling has become popular among researchers for analyzing large amounts of unstructured text data quickly and efficiently. The most common type of topic modeling algorithm is Latent Dirichlet allocation (LDA), which is frequently used in literature review studies. For instance, Das et al. (2016) used LDA to examine the research trends underlying the papers from Transportation Research Board Annual Meetings. Similarly, Sun and

Yin (2017) used LDA to discover themes and trends in transportation research. However, LDA has some limitations. For example, LDA cannot capture changes in the topic and word representations within documents as a function of covariate values (Kuhn, 2018). Also, LDA assumes no correlation between the topics, which is not always efficient. The LDA model's shortcomings can be overcome using Structural Topic Modeling (STM) technique. The STM models corpus structure by allowing document-level covariates (Das et al., 2017). The plate notations of the LDA and STM are shown in Figure 3.

Both LDA and STM have topic prevalence, language model, and topic content component. The STM model combines and extends three existing models, which are the correlated topic model, the Dirichlet multinomial regression topic model, and the sparse additive topic model (Blei and Lafferty, 2007; Roberts et al., 2013; Eisenstein et al., 2011). The core language model uses the correlated topic model by allowing correlations in the document-topic proportions with the use of the logistic normal distribution. For a model with K topics where k (i.e., $k \in \{1 \dots K\}$ is the index of the topics, the language model follows*:*

$$\eta \sim N(\mu, \Sigma) \qquad (1)$$

$$\theta_k = \frac{\exp(\eta_k)}{(\sum_i \exp(\eta_i)} \qquad (2)$$

Where, η_k is fixed to 0 for identification. A topic is sampled from a multinomial distribution $z \sim M(\theta)$ for each token within a document. Based on the sampled topic, a word is chosen from a distribution of words β_Z. Here μ and β are related to the document covariates.

The topic prevalence component of STM as shown in Figure 3 allows the expected document topic proportions to vary by covariates (X). The topic prevalence component models the mean vector of the logistic normal as a simple linear model such that $\mu_d = X_{d\gamma_d}, where\ \gamma$ is a regularization prior that prevents overfitting. On the other hand, the topical content component depends on the parameterization of the distribution over words as deviations in log space from a corpus-wide baseline *m*. For a mutually exclusive group of documents, the distribution over words is:

$$\beta_{d,k,v} \propto \exp\left(m_v + K_v^{.,k} + K_v^{Y,.} + K_v^{Y,k}\right) \qquad (3)$$

Where, *m* is the baseline log frequency of words, $K_v^{.,k}$ is the deviations due to each topic, $K_v^{Y,.}$ is the non-topic-specific deviations due to covariates Y, and $K_v^{Y,k}$ is the topic-specific covariate deviations.

2.2.2 Selection of number of topics

Identification of the number of topics is an important step in STM. There is no single correct method for selecting the number of topics (Kuhn, 2018). Measures such as semantic coherence, exclusivity, residuals, and held-out-likelihood can be used for estimating the number of topics. Kim et al. (2020) used these criteria to identify the optimum number of topics. Kuhn (2018) and

Das et al. (2017) identified several topics by studying the trade-offs between semantic coherence and exclusivity.

Semantic coherence was developed by Mimno et al. (2011). Semantic coherence defines how frequently individual words occur and pairs of distinct words co-occurrence (Kuhn, 2018). The value of semantic coherence is maximized when the most probable words in a topic frequently occur together. It can closely correspond to the human judgment of latent topics in documents (Mimno et al., 2011). Generally, the semantic coherence decreases as the number of topics increases.

The second criterion, called exclusivity, is that high-probability words under one topic should not overlap with high-probability words in other topics, which means that high-probability terms should be unique and exclusive to one topic only (Kim et al., 2020; Bischof and Airoldi 2012). Exclusivity has a positive relationship with the number of topics, meaning with the increase in topics, the exclusivity of the model increases.

The third factor is residual analysis, proposed by Taddy (2012). A topic model's theoretical multinomial error variance is one, and an error variance larger than one means that the true number of topics is greater than the current number of topics. For residual analysis, values close to one mean a better fit.

The held-out likelihood method estimates the probability of a word appearing in a document when those words have been retracted from the document during the estimation step (Wallach et al., 2009). In this method, some of the documents are removed from the estimation phase and used for validation purposes. The held-out likelihood values help to access the model's prediction performance. The held-out likelihood is "the log-likelihood calculated from the test set and conceptually analogous to a fit index of a confirmatory factor analysis model on a second data set after the model has been identified with an exploratory factor analysis model on the first data set" (Kim et al., 2020). Although these four diagnostics provide a good data-driven method to select the optimal number of topics from a set of documents, optimizing these parameters means improving the model fit without necessarily enhancing the substantive interpretation of the model (Cheng et al., 2009).

2.3 App Development

The research team developed a conceptual overview of the App (i.e., the concept of operations, *ConOps*), including the content's data sources, platforms for the showcasing, frequency of the update, and algorithm to grab and filter content from the internet. The App development process will follow the steps shown in Figure 4.

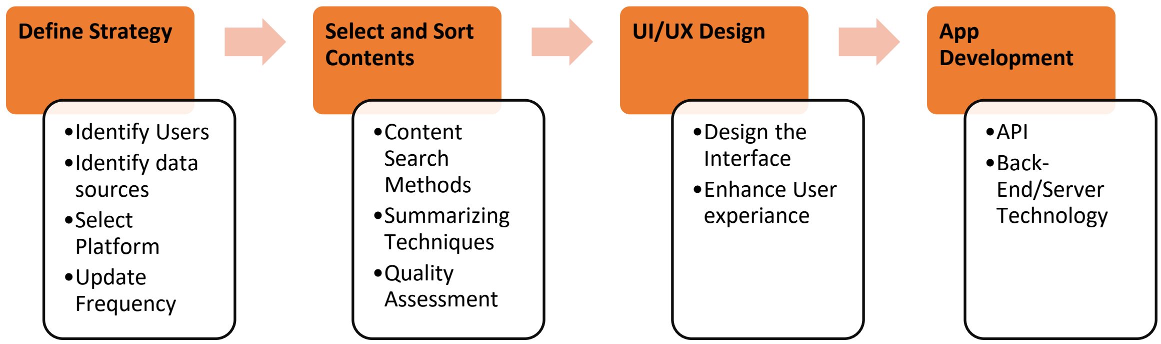

Figure 4: Steps in App development (Adopted from Invonto.com)

2.3.1 Strategy selection

The first phase of the App development was to define the strategies for selecting the data sources, user groups, a platform for showcasing App content, and the update frequency. As explained before, the App's main objective is to share the infrastructure durability and relevant research, news, and other relevant information for the DOTs, private/public sector professionals, researchers, and other interested stakeholders. The relevant contents were collected from major journals, magazines, websites, news media, social media (e.g., Twitter, Linkedin), Google search, Patent Search, and TRB TRID database. The information was provided via App downloaders. The admin App access allows DOTs, practicing engineers, consultants, and researchers to upload relevant information.

2.3.2 Stakeholder survey

The stakeholder needs and expectations from the App in gaining access to the infrastructure news are critical in designing the smartphone App. Moreover, it is essential to identify the stakeholder needs and expectations to ensure the effective selection of relevant content and presentation in the App in a user-friendly manner. Relevant information can be gathered by surveying the relevant stakeholders. A survey was developed and distributed among the stakeholders from state DOTs, university research centers, researchers, public/private sector practitioners, consultants, and student researchers in this study. The stakeholder survey's finding was used to determine stakeholder's needs (i.e., ConOps). Information on the App was updated weekly.

2.3.3 Content sorting

An algorithm was developed to collect up-to-date and relevant information on infrastructure durability and life-extension research. The algorithm contains keyword searches using a Python web crawler-based algorithm. The quality control method and summarization techniques, including the data fusion technique, were used in this stage. Before making the weekly updates/digest available to the public (e.g., DOTs, practicing engineers, consultants, and researchers), the research group performed quality control, and a process was developed for the long-term continuation of the App after completion of the project.

A Python language-based web-crawling algorithm has recently gained popularity in effectively extracting information from the web for subsequent analysis and decision-making. These web crawling algorithms can be used to effectively search selected websites to gather relevant and timely information in the field of infrastructure research and feed them into the App. Several Python -based web scraping frameworks are available (Kouzis-Loukas, 2016; Zheng, 2015; Chaulagain, 2017). Among these, Scrapy, beautiful soup, and selenium are the most popular ones. For the classification of collected information, a separate text classification/topic modeling algorithm was used (discussed in the previous section).

2.3.4 UI/UX design

The ultimate user benefit from the App depends on how well a user can search and navigate the App interface and find the desired contents. The UI/UX design aims to deliver a seamless experience to the users. In this App development phase, the research team explored different design aspects to make the interface interactive, intuitive, and user-friendly. Mockups and prototypes were developed in this stage to simulate the user experience and workflows expected from the App's final version.

2.3.5 App development

The backend or server technology (discussed in detail in Section 3.4) for the App was selected and an Application Programming Interface (API) to establish communication between the App and the database was defined. Figure 5 illustrates the App development process's data collection, integration, and user interface.

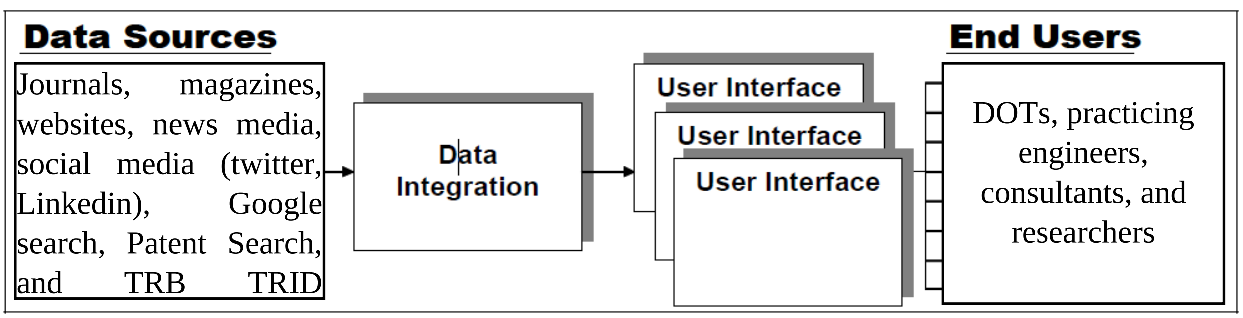

Figure 5: APP Data Flow Structure

2.3.6 App testing

The research team performed the App testing to ensure that the App was performing as expected and that the App was not producing any unintended output. App testing also ensures that the App is stable. The research team performed functional and performance tests. Unit tests and integration tests were performed for functional tests. A unit test consists of testing a single function of the App while the integration test consists of testing a large part or functionality of the App (Larga, 2020). In performance testing, issues such as App response to the user’s request, time to load the screen, App’s efficiency in using network bandwidth were investigated. The App architecture was modified during testing to address any issues readily visible from different testing activities.

Chapter 3. Data Analysis and App Development

3.1 Stakeholder Survey Results

3.1.1 Survey distribution

A survey of relevant stakeholders was conducted to understand user needs and expectations from the App. Invited stakeholders included state DOTs, university transportation research center researchers, public/private sector practitioners, consultants, and student researchers. A survey questionnaire was developed using the Qualtrics platform and distributed via email among stakeholders. The survey was distributed via an anonymous email link between August 29 and September 20, 2022, and 95 complete responses were collected. The survey included ten questions related to users' expectations from a smartphone App related to transportation infrastructure and durability. The survey questionnaire is provided in Appendix A. The following sections summarize the survey findings.

3.1.2 Survey participant's characteristics

The survey was distributed among stakeholders from state DOTs, university research centers, researchers, public/private sector practitioners, consultants, and student researchers. The distribution of the survey responses based on professional affiliation is shown in Figure 6. About 40% of the respondents were from university research centers, and 36% of the responses were from state DOT professionals, students, and professors.

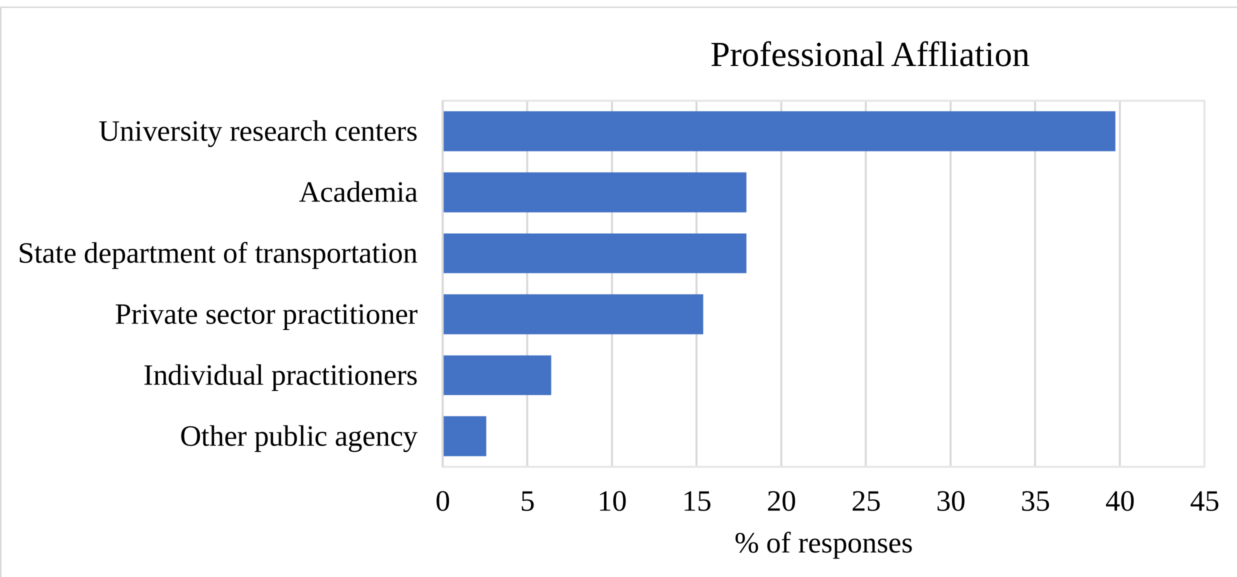

Figure 6: Professional Affiliation of the survey respondents. (Related Question: Q2, Appendix A)

The professional experience level of survey respondents in the transportation infrastructure durability and life extension-related field is shown in Figure 7. Most respondents had more than 10 years of professional experience (30% of the total respondents), 28% had 5 to 10 years of experience, and 23% had 1 to 5 years of experience.

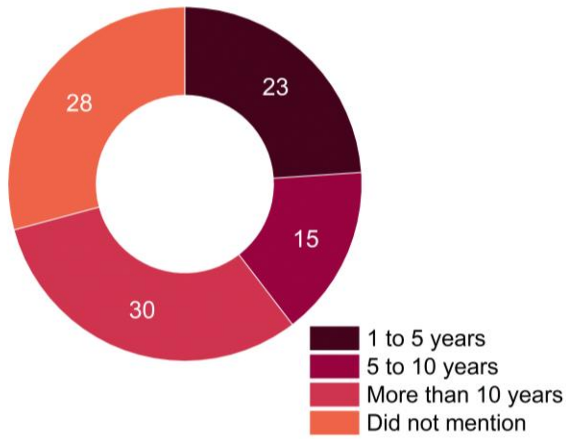

Figure 7: Respondent's experience in the infrastructure durability and life extension field. (Related Question: Q1, Appendix A)

3.1.3 User current practices

It is necessary to know how users currently access information from the internet and what type of information to design a smartphone App that properly serves the user's needs. Among the respondents, Google search and visiting dedicated websites were the main two sources for accessing information (56.5% of respondents, Figure 8). Social media and emailed/printed newsletters were also important sources for accessing the transportation infrastructure and durability-related news. Other sources of information include Transportation Research Board (TRB) news, library services, and journals. As shown in Figure 9, most respondents searched for research articles and reports (about 33% of respondents). In addition, survey respondents searched for best practices, the latest innovations, and seminar/webinar news.

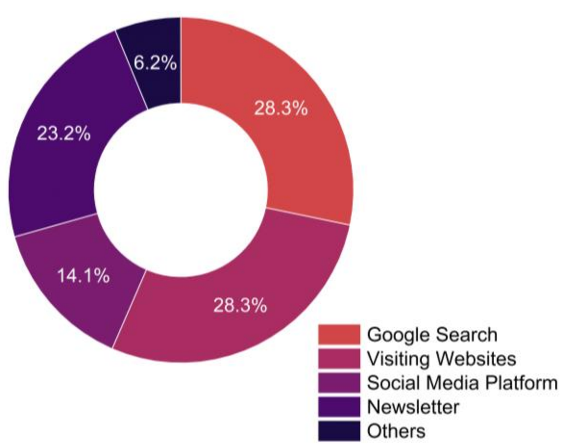

Figure 8: Sources of transportation infrastructure durability and life extension news and other information. (Related Question: Q3, Appendix A)

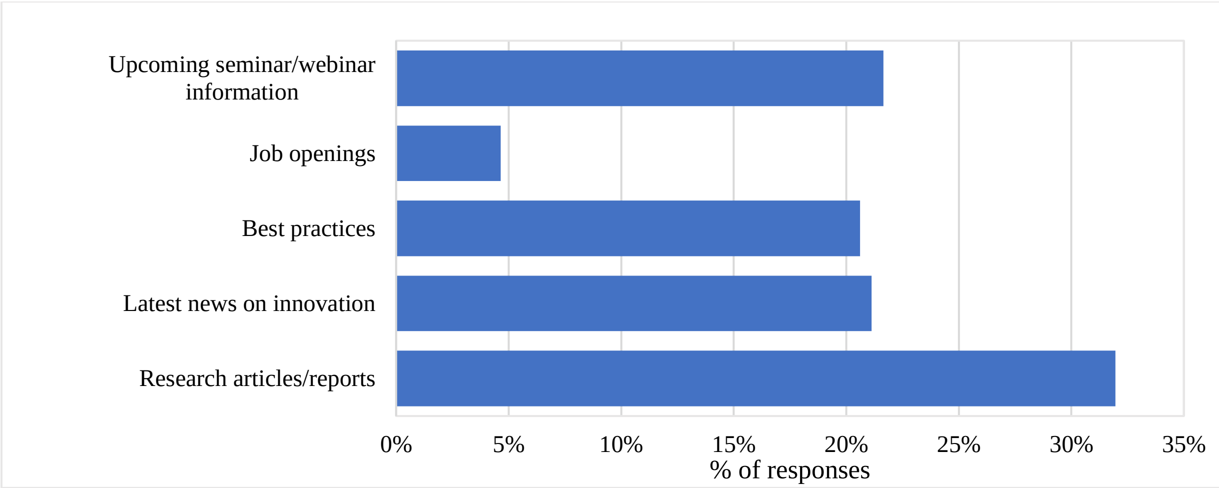

Figure 9: Types of information searched for related to transportation infrastructure durability and life extension. (Related Question: Q4, Appendix A)

Among the survey participants, 29% of the respondents visited the information sources at least once a week, and 29% visited two-three times a week (Figure 10). Only 6% of respondents visited the websites or journal pages daily, which suggests that weekly update is the most appropriate update frequency for the App.

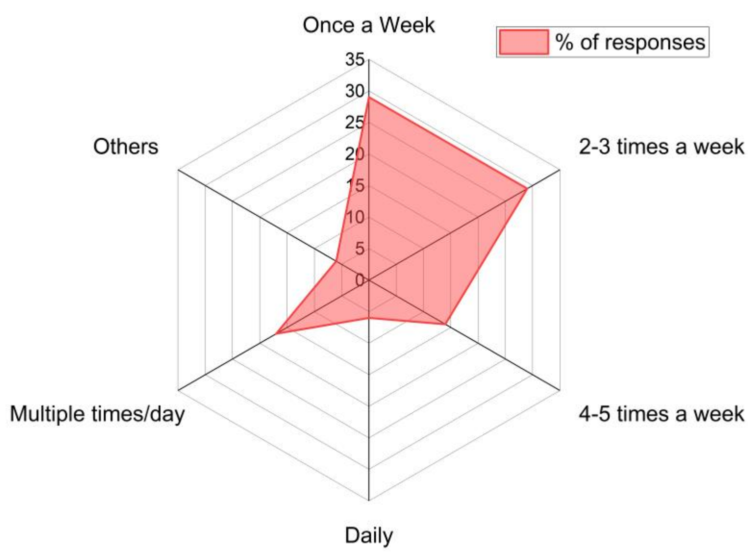

Figure 10: Frequency of visits to access transportation infrastructure durability and life-extension related news

3.1.4 User expectations

Survey participants shared their expectations regarding App update frequency, user's preference towards different app contents, and App features (Appendix A, survey questions 6 to 9). Figure 11 and Tables 1 and 2 summarize the survey responses.

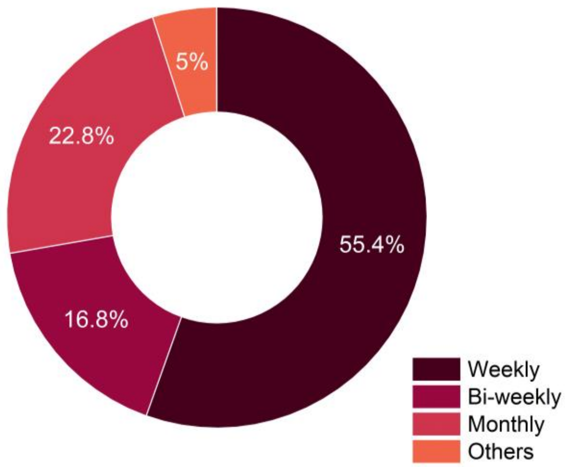

Figure 11: Expected frequency of updates from the APP (Related Question: Q6, Appendix A)

About 56% of the respondents expect a weekly update, 17% want a bi-weekly update, and 23% want a monthly update (Figure 11). Respondents were asked to rank different components and features in designing the App, and findings are summarized in Tables 1 and 2 (Survey questions 7 and 8, Appendix A). The expected highest-ranked content was *Latest news and advancements in the field*, followed by a *Summary of the latest research projects and reports* and *Educational and professional training opportunities.* Among the App's features, *Summary of the content* got the highest preference. *Visual aesthetics* and *ease of app navigability* were other two important expected features.

Table 1: Rank of the smartphone App contents that are more important to the user (Related Question: Q7, Appendix A)

#	Field	Mean	Std. Deviation
1	Summary of latest research projects and reports	7.9	2.2
2	Latest news and advancements in the field	8.0	1.9
3	Educational and professional training opportunities	6.1	2.3
4	Job opportunities	4.0	2.9

Table 2: Rank the smartphone App features based on the survey responses (Related Question: 8, Appendix A)

#	Field	Mean	Std. Deviation
1	Frequency of the update	6.45	2.47
2	Summary of the content	8.32	2
3	Ease of App navigability	8.3	2.01
4	Visual aesthetics	7.14	2.44

The findings from the survey conducted in this study were used in App design which involves developing a conceptual overview of the App, including the content's data sources, platforms for the showcasing, frequency of the update, and algorithm to grab and filter web content. The weekly update frequency for the App was the most favorable based on the user's current practice and expectations. Research articles, projects, and the latest news on recent innovations were the most searched among the transportation infrastructure durability and life-extension related practitioners. In addition, the App design was focused more on providing content that efficiently summarizes the latest news, research projects, and educational and training opportunities. The App design also focused on visual aesthetics and ease of navigability for a better user experience.

3.2 Database Selection and Literature Collection

3.2.1 Collected literature

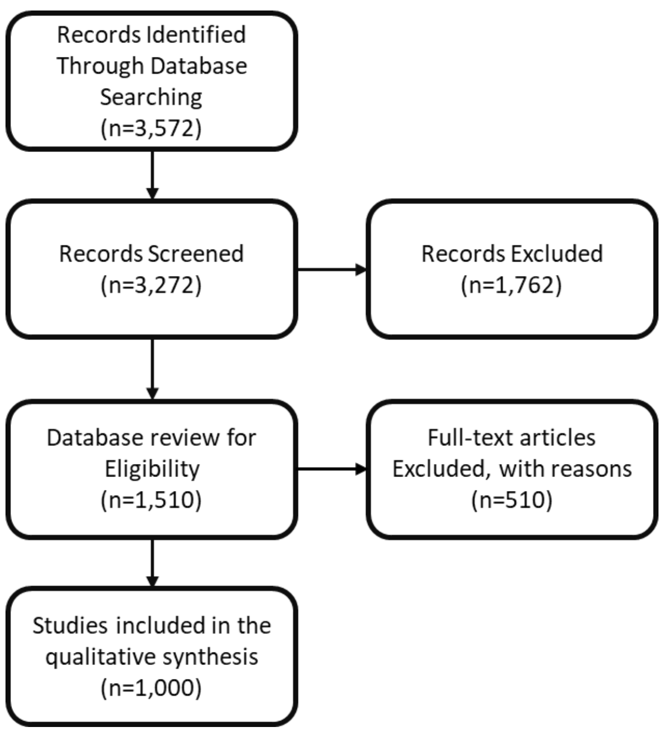

Figure 12: The flow chart of study selection and screening process

Relevant literature was collected from three databases using the search strings provided in section 2.1.2. Collected studies were then screened based on the exclusion and inclusion criteria presented in Section 2.1.1 and the PRISMA flow diagram in Figure 2. The literature screening result is summarized in Figure 12. While the literature search against the databases resulted in approximately 3,572 papers, around 300 were removed as they were not in the English language. After that, the remaining 3,272 papers were checked for duplicates, journal impact factors, missing data, and publication types to exclude them from the final database. In this process, 1,762 papers were excluded (Figure 12). The titles of the remaining 1,510 papers were carefully reviewed to assess their eligibility for the current study. An additional 510 papers were removed in this process due to a lack of relevance to the durability of transportation infrastructure and life extension. As a result of the screening process, 1,000 studies are selected in this paper related to transportation infrastructure (i.e., pavement, bridge, tunnel, and airfield), durability, and life extension and subsequently analyzed in the qualitative synthesis.

Table 3: Distribution of the literature based on the type of bibliographical reference.

Publication Type	**Year**		
	2020	**2021**	**2022**
Reports	281	235	61
Conference Papers	2	5	1
Journal Articles	892	1133	944
Book Chapters	4	7	7

Table 3 represents the year of publication and publication categories of the included studies in the review process. The number of papers on transportation infrastructure durability and life-extension has increased over the years. There was 24% more literature in 2021 compared to 2020, and until June 2022, there are already 1,013 studies related to the topic, almost 74% of the publications in 2021. The majority of the collected documents were journal articles.

3.3 Research Trend Analysis

This section provides the result of data collection and data screening steps. The number of topic selections and topic modeling are also discussed.

3.3.1 Topic selection

Selection of the optimum number of topics is required before running the final topic modeling algorithm. The number of topics was selected based on the criteria presented in Section 2.2.2. At first, different STM models with varying topics were developed, and the held-out likelihood, residuals, semantic coherence, and lower bound were calculated. For this purpose, the number of topics was kept between 10 to 30. The resulting graphs are shown in Figure 13. Four indices did not show unanimous support for the number of topics. Based on the held-out likelihood, the optimum number of topics was around 25. The residual value showed the lowest value, around 30 topics. In terms of semantic coherence, the score decreased after 12 topics. According to Roberts et al. (2017), 3 to 10 topics are optimum for a short corpus (e.g., around a few hundred documents) with a strong and specific subject matter. Thus, the number of topics was set to 12 for this study to avoid an excessive number of topics with overlapping words. This balances the semantic coherence and held-out likelihood values. A lower number of topics helps to further investigate the state-of-the-art within the topics.

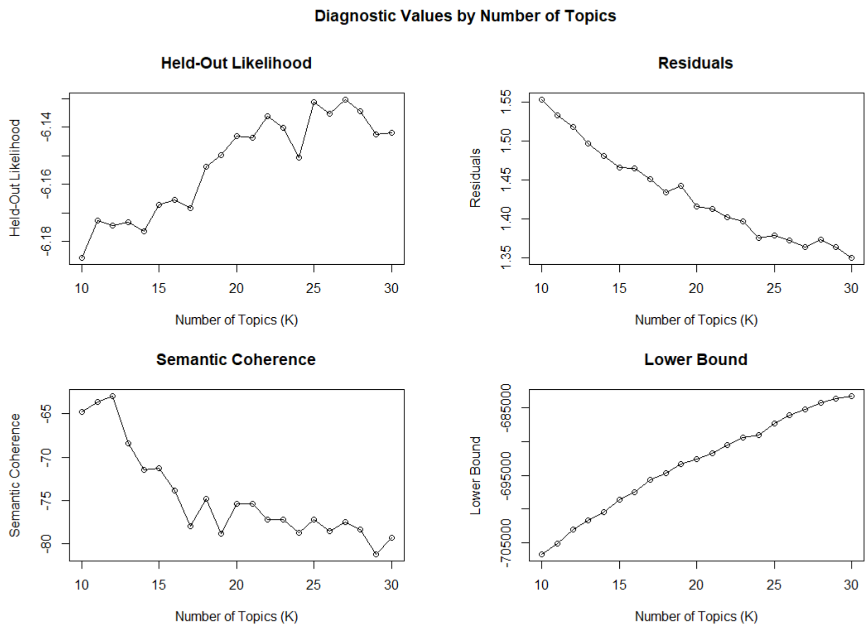

Figure 13: Selection of number of topics

3.3.2 Results from Topic Modeling

The final STM model was run with 12 topics. The STM model provides specific words that are linked to the resulting topics. The STM model did not provide explicit definitions or labels for the topics. The label for each topic was determined based on the frequency of different words under each topic and using the subject knowledge. This entails inspection of each topic one by one and examining a small collection of words that appear under that topic based on probability, lift, and FREX statistics (Kuhn, 2018). For example, the words data, model, prediction, forecasting, and Markov were linked to topic 1. So, the label of the topic was given as “data use and statistical models”.

The simplest way to assign labels to a topic would be to select the words with the highest probability of appearing under that topic. However, this could lead to incorrect labeling as certain words, such as pavement and roads, appear as high-probability words under many topics. These common words were frequent in the database but added relatively little value to the topic labeling. For this reason, other metrics such as FREX, Lift, and Score were used to help identify a topic's correct label. Lift score refers to the probability of word occurrence in a topic divided by the probability of the word occurring across the whole corpus (Kuhn, 2018). The lift metric selects the words that are much more common in a topic than the entire dataset. FREX statistics, proposed by Bischof and Airoldi (2012) calculate the ratio of word frequency conditional on a topic to word-topic exclusivity (Kuhn, 2018). FREX statistics measure the exclusivity of a word by balancing the word frequency.

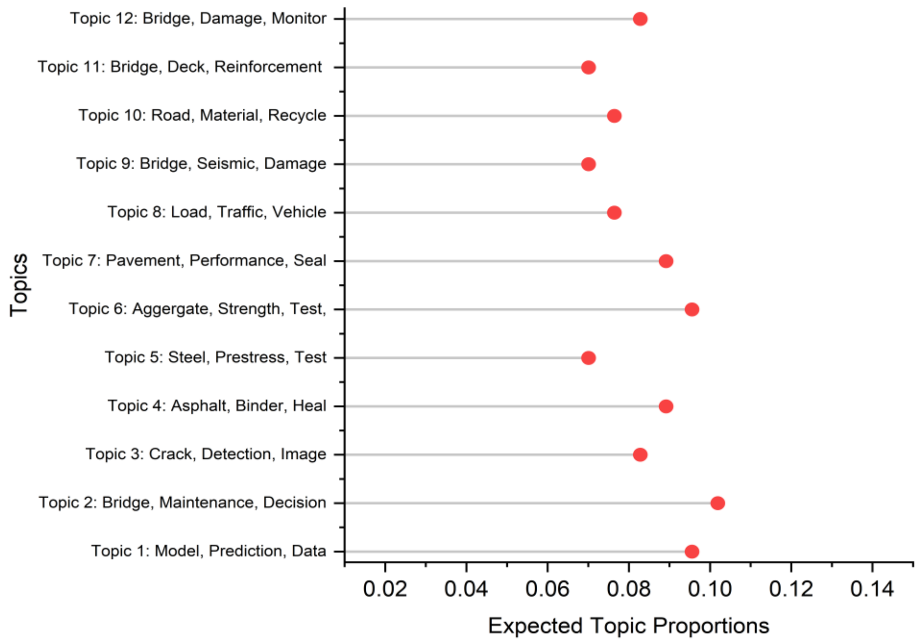

Figure 14: Extracted topics from the STM model and expected topic proportions

As shown in Figure 3, the STM model can incorporate covariates into the modeling framework. The covariates affect the topic prevalence or topic content latent variables with the document-level information (Das et al., 2017). Covariates for topical prevalence models show how the observed metadata affects the frequency with which a topic is discussed, and the covariates for topical content models show how the observed metadata affects the word rate use within a given topic. In this study, the publication year and name of the journals were used as the covariates in the topical prevalence portion of the model. Before running the STM model, text preprocessing was performed, such as stop words removal, stemming (i.e., reducing words to the root form), number, and special character removal. The R package named "STM" was used to estimate structural topic modeling and result visualization (Roberts et al., 2013). The 12 topics from the STM model, along with the top three keywords and expected topic proportions, are shown in Figure 14. Topic 2 had the highest expected topic proportions of 0.10. The top three words under this topic were "Bridge", "Maintenance", and "Decision". The identified topics and their associated words showed that high-frequency topics were related to project management, decision support systems, use of data and AI, materials, and traffic.

Table 4 shows the top 12 topics sorted by the expected topic proportions from the final STM model and the top five words under each topic based on the probability, lift, score, and FREX statistics.

As there is no single correct statistic to select words for labeling the topics, all four metrics (i.e., probability, FREX, lift, and score) were used for the topic labeling. As presented in Table 2, topic 1 focused on the data-based statistical model (e.g., predict, Markov, data, and stochastic), while topic 3 was related to machine learning and image processing in pavement crack detection and bridge damage identification. Compared to these method-oriented topics, topics 6, 4, and 14 were related to material research for transportation infrastructure, and topics 2 and 7 highlighted the performance monitoring and decisions support system on infrastructure maintenance.

Table 4: Top 12 topics and labels

Topic ID	Topic Label	Criteria	Word 1	Word 2	Word 3	Word 4	Word 5
2	Bridge Maintenance Decision System	Probability	bridge	maintenance	system	cost	manage
		FREX	decision	asset	manage	plan	maintain
		Lift	interest	maker	asset	allocation	fund
		Score	interest	asset	manage	bridge	decision
1	Data and statistical models	Probability	model	predict	method	propose	use
		FREX	predict	data	methodology	approach	propose
		Lift	budget	Markov	update	stochastic	model
		Score	budget	model	predict	Markov	update
6	Concrete and Aggregate Properties	Probability	aggregate	use	material	pavement	concrete
		FREX	aggregate	cement	cure	thaw	skid
		Lift	abrasive	texture	unconfine	skid	ash
		Score	texture	aggregate	cement	skid	ash
7	Pavement performance monitoring and treatments	Probability	pavement	perform	overlay	asphalt	crack
		FREX	overlay	treatment	pavement	seal	section
		Lift	FDR[1]	slurry	AASHTO[2]	reclaim	LTPP[3]
		Score	pavement	treatment	overlay	seal	asphalt
4	Asphalt Binder Performance and Modeling	Probability	asphalt	mixture	perform	test	binder
		FREX	rap	heal	binder	mixture	asphalt
		Lift	DSR[4]	bitumen	heal	rheometer	viscos
		Score	asphalt	mixture	binder	rap	heal
12		Probability	bridge	damage	structure	cable	monitor
		FREX	cable	monitor	frequency	vibrate	stay

	Bridge Damage Detection	Lift	modal	slide	SHM[5]	stay	cable
		Score	slide	bridge	cable	vibrate	modal
3	Crack Detection Using AI	Probability	crack	detect	method	image	data
		FREX	image	neural	learn	detect	GPR
		Lift	pixel	convolution	recall	recognition	CNN
		Score	pixel	detect	image	CNN	crack
10	Renewable and Sustainable Pavement Materials	Probability	road	pavement	use	construct	material
		FREX	recycle	road	mark	retroreflect	review
		Lift	mark	retroreflect	topic	driver	recycle
		Score	retroreflect	recycle	mark	pavement	road
8	Effect of Traffic on Pavements and Bridges	Probability	load	traffic	vehicle	test	truck
		FREX	truck	load	deflect	traffic	wave
		Lift	wave	passage	weigh	back	axle
		Score	wave	load	wind	truck	vehicle
9	Vulnerability to Natural Hazards	Probability	bridge	seismic	damage	flood	scour
		FREX	seismic	pier	collapse	earthquake	event
		Lift	hydraulic	collapse	flood	scour	fragile
		Score	hydraulic	seismic	pier	bridge	earthquake
5	Bridge Girder	Probability	steel	girder	test	beam	bridge
		FREX	prestress	beam	girder	steel	shear
		Lift	strand	prestress	web	anchor	flexure
		Score	strand	girder	prestress	steel	bar
11	UHPC, FRP, Glass Fiber Reinforced	Probability	bridge	deck	concrete	reinforce	use
		FREX	deck	reinforce	replace	detail	FRP
		Lift	FRP[6]	closure	deck	official	wall
		Score	FRP	bridge	deck	UHPC[7]	reinforce

1 FDR= Full-depth Reclamation; 2 AASHTO= American Association of State Highway and Transportation Officials; 3 LTPP= Long-term Pavement Performance; 4 DSR= dynamic shear rheometer; 5 SHM= Structural Health monitoring; 6 FRP= Fiber-reinforced polymer; 7 UHPC= Ultra-high-performance concrete.

3.4 App Development

App architecture development is the most important step towards building a well-structured and functioning APP. The mobile App architecture is a roadmap and detailed guide for creating and operating the App. It comprises structural elements and interfaces, several layers, connections between design elements, and the general program style. The multilayer architecture of the App consists of four layers, (i) Presentation layer, (ii) Business layer, (iii) Service layer, and (iv) Data layer, as shown in Figure 15. The detailed description of the layers is discussed in the following subsections. App architecture was followed for the complete application process from end to end.

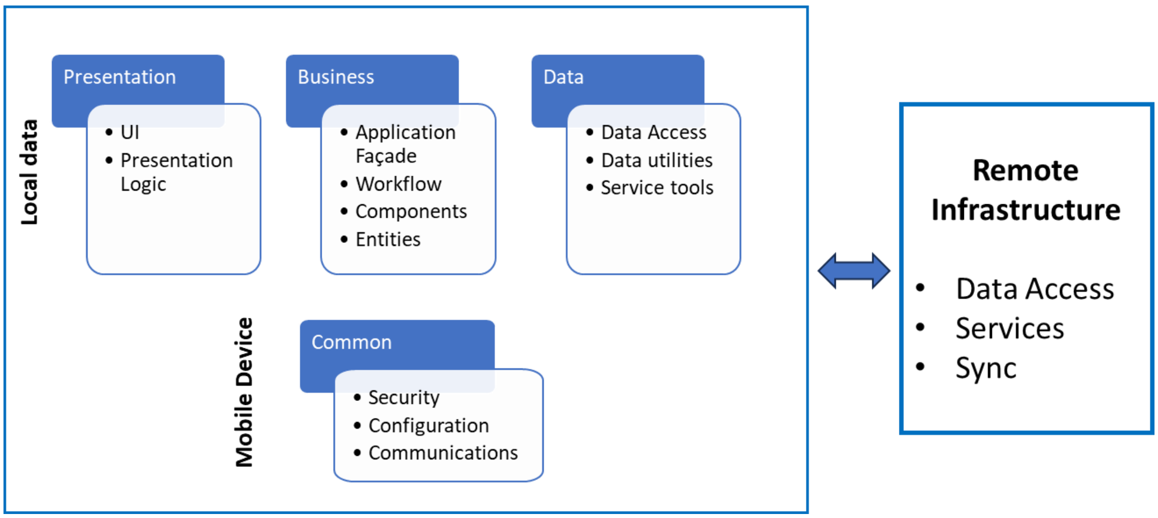

Figure 15: Mobile App architecture design components (Source: os-system.com)

Presentation Layer – The presentation layer contains the user interface (UI) components and UI process components. At a high level, this layer defines the App's presentation logic and how the product will look.

Business Layer – This layer focuses on demonstrating the business identity to the user, consisting of workflows, business entities, and components.

Service Layer - A Service Layer defines an App's boundary and available operations from the interfacing client layer's perspective. It encapsulates the App's business logic, controlling transactions and coordinating responses in implementing operations.

Data Layer – The data layer works on data-related components including access components, helpers, utilities, and service agents.

3.4.1 Frontend development

We designed the wireframes using Figma software per the App requirement for frontend development. The designed wireframes were developed with flutter and Dart code using BLoC pattern. For every functionality, flutter for building cross-platform applications was used. All the logics were written in the frontend, and API calls were used to fetch data from the backend.

3.4.2 Backend development

This project used backend methodology to create and maintain data, apply business logic, and expose it through RESTful APIs. For this, the spring boot framework was used to build all the APIs for every functionality. All the user data would be fetched into the database, which was designed using the MySQL workbench Server for this App.

3.4.3 Database

The database was designed using MySQL workbench for storing the user details. Once the user fills in all the details in the App, information was stored in the database through the API connection. The database table was designed with 'username', 'email', 'mobile number', 'password', 're-enter password' for the first-time user. For the existing user, 'user', 'password', were created, which were designed to fetch and get the user details from the database with the API connection.

3.4.4 Security

For security, the App used JWT authentications using Bearer Token. Swagger was for API documentation, and JSON was used for the API data exchange format. A JSON web token (JWT), a JSON Object, was used to transfer information over the web securely. An authentication system can be used for information exchange as well. The token has a header, payload, and signature.

3.4.5 Server

The App was deployed in the Heroku-based cloud Platform as a Service (PaaS) for the server. The steps followed in deploying the App were-

Step 1: Create a Heroku App and Heroku Postgres Database
Step 2: Provision the Heroku Connect Add-on
Step 3: Install the Heroku Connect CLI Plugin
Step 4: Add User to Heroku Connect
Step 5: Configure the Connection to Heroku Postgres
Step 6: Configure the Connection to Salesforce
Step 7: Check the Connection State
Step 8: Set Up Mappings

3.4.6 App functionality

The user App consists of four pages: Introduction page, Sign-in page, Skip-page, and Homepage, as shown in Figure 16. The first page is the Introduction page, where the user can see the welcome message/information and "Let's Go" button, which takes to the next page. The next page consists of the Login page, where users can Sign up or give their credentials to Login to the App. The next page consists of the user details page with all the related functions, i.e., Forgot Password, Sign Up functionality pages. Once the user sign in, they can see the main pages with all the functionalities. All the modules of the App are discussed below.

Once the user installs the App, the user can log in or create a new account by swiping the page (Figure 17) and seeing the App logo and Let's Go push button. The user can navigate to the user details pages, shown in Figure 18. Once the user Sign-in, the following tabs are displayed: (i) Updates, (ii) News, (iii) Research, (iii) Opportunities. These sub-modules are discussed below:

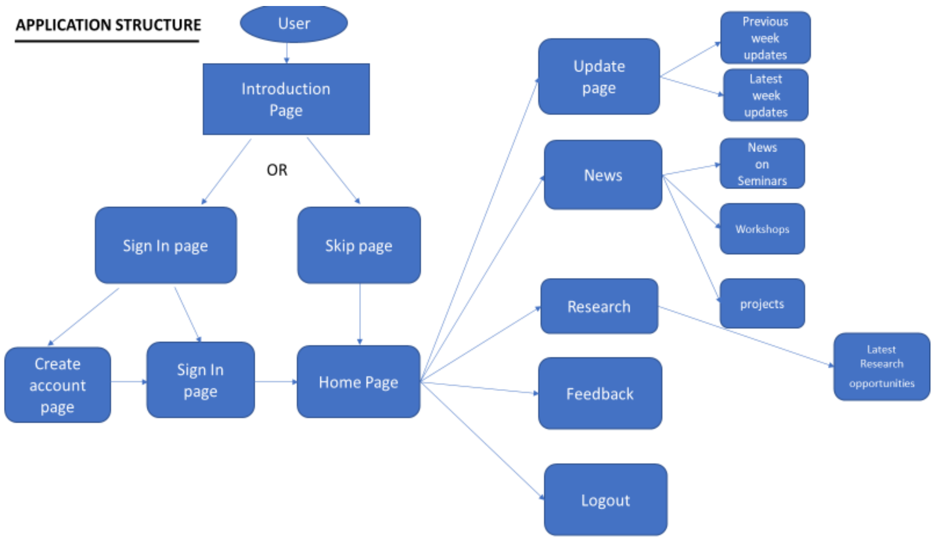

Figure 16: Application flow

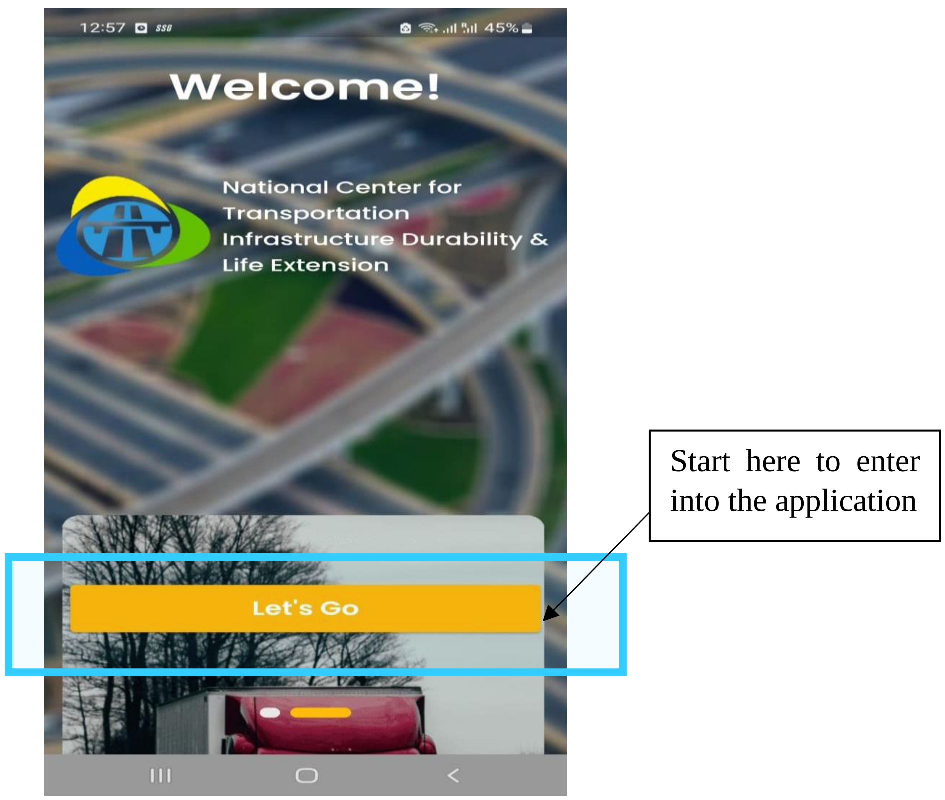

Figure 17: User App main page

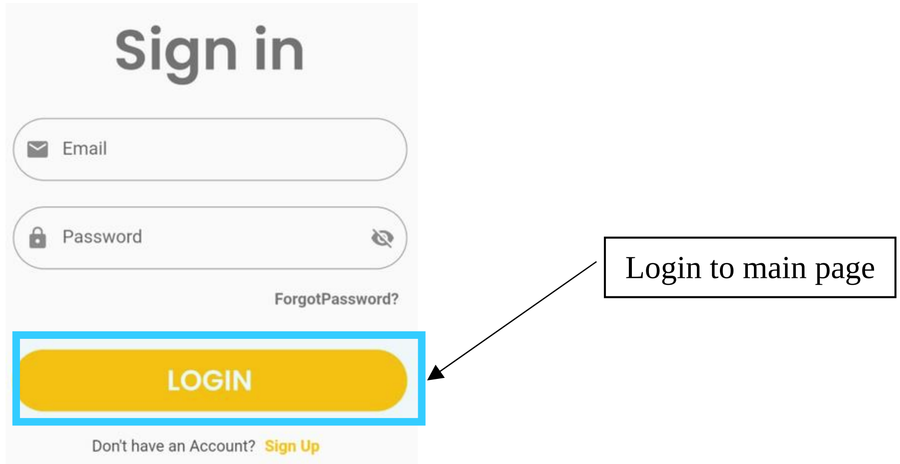

Figure 18:Login page

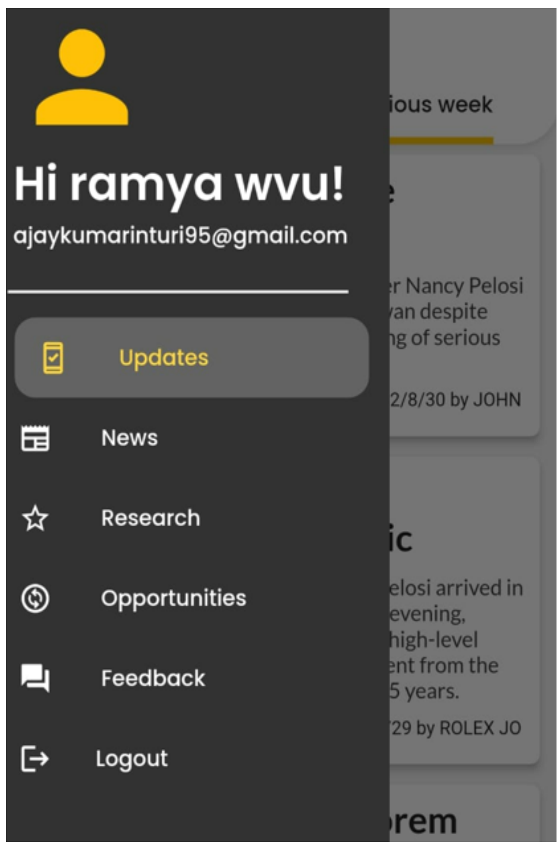

Figure 19: Main modules page

The “Updates” page consists of the most recent week’s updates and precious weeks' updates related to transportation infrastructure research. Once users click the “Updates” page, they can view “This Week” and “Previous Week” updates tab. The “This Week” tab consists of the updates from the most recent week. On the other hand, the “Previous Week” tab consists of the contents from the past four weeks. “News” page has three sub functionalities- Seminars, Workshops, and Projects. The “Seminars” page shows all the related information regarding upcoming seminars in the transportation infrastructure fields. Similarly, all the forthcoming workshop topics and schedules are displayed on the “Workshops” page. The “Projects” page shows the ongoing and upcoming project information. “Research” page shows the research findings and topics related to transportation infrastructure durability and life-extension. The “Opportunities” page lists job opportunities related to transportation infrastructure fields.

Chapter 4. Conclusion

Increasing the service life and durability of transportation infrastructure has become a critical challenge in developing and maintaining effective and reliable transportation systems under different traffic and environmental factors. New technologies, materials, and monitoring systems for transportation infrastructure have been developed to extend the service life of roads, bridges, tunnels, etc. This study developed a smartphone App to disseminate research articles/reports from studies by diverse organizations and relevant professional activities such as seminars, webinars, and workforce development opportunities. This study selected the Android platform to develop an App that compiles advancements in research and practices related to transportation infrastructure durability, life-extension, and other professional development opportunities. This study demonstrated the application of the STM technique to reviewing published studies between January 2020 and June 2022 relevant to transportation infrastructure durability and life-extension, and presented a holistic overview of state-of-the-art research and practice trends of service life extension and durability of transportation infrastructure.

The literature (i.e., reports, journal articles, conference papers, and book chapters) was collected and processed by applying the PRISMA framework. Using the STM model, 12 topics were identified based on the semantic coherence of the topics. The labels of the topics were identified using the words associated with the topics based on probability, FREX, and lift statistics, and applying the subject-matter expertise/knowledge of the authors. Most studied topics related to using AI/ML techniques and data, research on innovative materials, performance monitoring, MR&R activity selection, optimization, and the effect of natural hazards and climate change. Further analysis and discussion on these topics identified the state-of-the-art status and future research needs. Topic modeling showed advancements in AI/ML applications in infrastructure asset management and related fields in the last three years. Deep learning models gained popularity over traditional ML models in pavement crack detection and classification problems. Few studies explored climate change's effects on infrastructure deterioration and associated MR&R activities. Future research efforts should focus on the impact of the environment and extreme weather conditions on the infrastructure durability and life-extension and the development of cost-effective prevention and mitigation measures.

References

Berman, M., Costello, P., & Ballard, L. (2008). Developing a One-Stop Shop for Public/Specialized Transportation Information in Montana (No. FHWA/MT-08-006/8188). Montana. Dept. of Transportation. Research Programs.

Chaulagain, R. S., Pandey, S., Basnet, S. R., & Shakya, S. (2017, November). Cloud based web scraping for big data applications. In 2017 IEEE International Conference on Smart Cloud (SmartCloud) (pp. 138-143). IEEE.

FAST Act Summary. https://www.fhwa.dot.gov/fastact/summary.cfm. Accessed March 30, 2021

ASCE Infrastructure Report Card (2021). https://infrastructurereportcard.org/. Accessed March 30, 2021.

Kouzis-Loukas, D. (2016). Learning scrapy. Packt Publishing Ltd.

Larga, F. E. D. S. V. (2020). A mobile tour guide app for sustainable tourism (Doctoral dissertation).

Minaee, Shervin, et al. "Deep Learning--based Text Classification: A Comprehensive Review." ACM Computing Surveys (CSUR) 54.3 (2021): 1-40.

Mobile App Development Process: A Step-by-Step Guide [Updated for 2021]. (2021). https://www.invonto.com/insights/mobile-app-development-process/. Accessed March 30, 2021.

Datacamp. Making Web Crawlers Using Scrapy for Python (2021). https://www.datacamp.com/community/tutorials/making-web-crawlers-scrapy-python. Accessed March 30, 2021.

TriDurLE. https://s3.wp.wsu.edu/uploads/sites/2442/2021/03/TriDurLE-Brochure.pdf. Accessed March 30, 2021

Zheng, C., He, G., & Peng, Z. (2015). A Study of Web Information Extraction Technology Based on Beautiful Soup. JCP, 10(6), 381-387.

Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health information & libraries journal. 2009 Jun;26(2):91-108.

Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British journal of management. 2003 Sep;14(3):207-22.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology. 2009 Oct 1;62(10):e1-34.

Pahlevan-Sharif S, Mura P, Wijesinghe SN. A systematic review of systematic reviews in tourism. Journal of Hospitality and Tourism Management. 2019 Jun 1;39:158-65.

Bai X, Zhang X, Li KX, Zhou Y, Yuen KF. Research topics and trends in the maritime transport: A structural topic model. Transport Policy. 2021 Mar 1;102:11-24.

Das S, Sun X, Dutta A. Text mining and topic modeling of compendiums of papers from transportation research board annual meetings. Transportation Research Record. 2016;2552(1):48-56.

Sun L, Yin Y. Discovering themes and trends in transportation research using topic modeling. Transportation Research Part C: Emerging Technologies. 2017 Apr 1;77:49-66.

Kuhn KD. Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transportation Research Part C: Emerging Technologies. 2018 Feb 1;87:105-22.

Das S, Dixon K, Sun X, Dutta A, Zupancich M. Trends in transportation research: Exploring content analysis in topics. Transportation Research Record. 2017;2614(1):27-38.

Blei DM, Lafferty JD. A correlated topic model of science. The annals of applied statistics. 2007 Jun;1(1):17-35.

Roberts ME, Stewart BM, Tingley D, Airoldi EM. The structural topic model and applied social science. InAdvances in neural information processing systems workshop on topic models: computation, application, and evaluation 2013 Dec 10 (Vol. 4, pp. 1-20).

Eisenstein J, Ahmed A, Xing EP. Sparse additive generative models of text. InProceedings of the 28th international conference on machine learning (ICML-11) 2011 (pp. 1041-1048).

Bischof J, Airoldi EM. Summarizing topical content with word frequency and exclusivity. InProceedings of the 29th International Conference on Machine Learning (ICML-12) 2012 (pp. 201-208).

Appendix A: Survey Questions

Q1. Are you currently involved in transportation infrastructure durability and life extension-related project planning, implementation, or research?

- ○ Yes (1)
- ○ No (2)

If yes, how long (in years) have you been working in transportation infrastructure durability and life extension?

____

Q2. Please select one of the following options that reflect your affiliation.

- ○ State Department of Transportation (1)
- ○ Other public agency (2)
- ○ Private sector practitioner (3)
- ○ University research centers (4)
- ○ Individual practitioners (5)
- ○ Others (6) ____

Job Title: ____

Email Address: ____

Q3. How do you access the latest news and the most up-to-date information about the research products, project reports/findings, and best practices related to transportation infrastructure durability and life extension? Select all that apply.

- ○ Google search (1)
- ○ Visiting websites (such as journals or organizations) (2)
- ○ Social media platform (3)
- ○ Newsletter (email/printed) (4)
- ○ Others (Please specify): (5) ____

Q4. Which types of information related to transportation infrastructure durability and life extension were you seeking during your most recent visit to the preferred source(s) (based on your response to the last question)? Select all that apply.

- ○ Research articles/reports (1)
- ○ Latest news on innovation (2)
- ○ Best practices (3)
- ○ Job openings (4)
- ○ Upcoming seminar/webinar information (5)
- ○ Others (Please specify): (6) ____

Q5. How frequently do you visit preferred information source(s)?

- ◯ Multiple times a day (1)
- ◯ Once a day (2)
- ◯ 4-5 times a week (3)
- ◯ 2-3 times a week (4)
- ◯ Once a week (5)
- ◯ Others (Please specify): (6) ____

Q6. What is your update frequency expectation for the smartphone App on transportation infrastructure durability and life extension?

- ◯ Weekly (1)
- ◯ Bi-weekly (2)
- ◯ Monthly (3)
- ◯ Others (Please Specify): (4) ____

Q7. Rank the smartphone App contents that are more important to you (10 for highest priority and 1 for lowest priority).

0 1 2 3 4 5 6 7 8 9 10

Summary of latest research projects and reports ()	
Latest news and advancements in the field ()	
Educational and professional training opportunities ()	
Job opportunities ()	
Others (specify): ()	
Others (specify): ()	

Q8. Based on your preference, please rank the smartphone App features (10 for highest priority and 1 for lowest priority)

0 1 2 3 4 5 6 7 8 9 10

Frequency of the update ()	
Summary of the content ()	
Ease of App navigability ()	
Visual aesthetics ()	
Others (Please Specify): ()	

Q9. What are your additional preferred deliverable types posted on social media or other online sources? (10 for highest priority and 1 for lowest priority).

0 10 20 30 40 50 60 70 80 90 100

Newsletter ()	
One Page Infographics ()	
Email Digest ()	
Others (Please Specify) ()	

Q10. Please share with us if you have any other comments and opinions on the function and scope of the smartphone App for transportation infrastructure durability and life extension.

____

Appendix B: User Guide

App Installation

The infrastructure App was developed for the Android platform and can be installed at any Android device, including Tablets or Smartphones. In addition to the main App, the developer team also created an admin App to update the contents of the main App. The admin App can only be accessed by specific people with permission. Both APK files of the main App and the admin App can be found at the following address:

Google Drive:

https://drive.google.com/drive/folders/1i25IVqUze9nJRbLgpyBQQjofw7GB7idO?usp=sharing

Or on the GitHub at:

https://github.com/tanvirshafad19/Tridurle-app

NOTE: Due to the sensitive nature of the API key, the user must be given permission to access the GitHub (i.e., submit a request to the developer).

After accessing the above links, there are two choices. In the first method, the APK files can be downloaded onto a computer, and then they can be moved to the desired device via USB cable. For the second method, the link could be accessed via phone and the APK files can be downloaded directly. The following article can be used as a guideline for this process:

https://www.lifewire.com/install-apk-on-android-4177185

Note that the easiest way is to download the APK onto the computer, connect the phone, after enabling USB debugging (https://www.lifewire.com/enable-usb-debugging-android-4690927) find the APK in files, and install it.

The user App consists of two main pages. The first page is the Sign in or Sign up page, depending on the user. The second page, Home Page contains the main content of the information, including the updated information, previous week information, research, seminars, workshops, and job opportunities. The other two modules consist of the feedback page and the logout page.

Upon loading the App for the first time, it will request the user to sign up if the user is a first-time user. If they want to create an account in the App, they can enter their information and sign up. If they do not want to create an account, they can skip the sign up, and the App will take the user to the main home page. For the returning user, they can either sign in using the credentials provided during the sign-up process, or they can also skip and go to the Home page directly.

Project Code:

To allow the next update or maintain the App for future use, the developer has shared all the App codes in a google drive folder. The project codes can be accessed from the following links. For security reasons, accessing the below links requires permission. If you want to access the codes for this project, please contact the developer at the following email address: Ajay Inturi (ai00012@mix.wvu.edu).

https://drive.google.com/drive/folders/1gwnJ4GF0oSx93ibo1DloazXKCG1H7s9t?usp=sharing