

ACUTE - Accessibility and Connectivity Knowledge Hub for Urban Transformation in Europe

WP4 – Strategic Support for DUT

D4.1 Report on synthesized results for the development of future DUT programme

Start date: 01.11.2022

End date: 31.10.2024

Authors: Irina Arhipova
Līga Paura
Nikolajs Bumanis
(D1.4. Enrica Papa and Sabina Cioboata)
(D2.2.&D2.3. Göran Smith, Christina Lindkvist, Isak Vencu Öhrlund & David Mowitz)

Version: Final Public

Project Partners

Organisation	Country
University of Natural Resources and Life Sciences, Vienna BOKU	AUSTRIA
Université Gustav Eiffel UEiffel	FRANCE
Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement Cerema	FRANCE
Latvia University of Life Sciences and Technologies LBTU	LATVIA
University of Latvia LU	LATVIA
Research Institutes of Sweden RISE	SWEDEN
University of Westminster UoW	UNITED KINGDOM
Malmö University MAU	SWEDEN
Grazer Energieagentur GmbH, Graz Energy Agency GEA	AUSTRIA
VTI/Sweden's national centre for research and education on public transport K2	SWEDEN
Power Circle PC	SWEDEN
University of Innsbruck UIBK	AUSTRIA

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission and JPI Urban Europe is not responsible for any use that may be made of the information it contains.

Table of Contents

Executive Summary	3
Introduction	4
1. Textual analysis	5
1.1. Pre-processing	7
1.2. Word frequency analysis	9
1.2.1. Analysis of project proposal's report texts	10
1.2.2. Analysis of the project's 1 st & 2 nd year report texts	11
1.2.3. Analysis of project final report texts	13
1.3. Similarity analysis	14
1.3.1. Relationships of project proposal's report text	15
1.3.2. Relationships of project 1 st & 2 nd year report text	19
1.3.3. Relationships of project's final report texts	22
1.4. Unique terms identification and text gap analysis	24
1.4.1 Uniqueness in project proposal's report text	26
1.4.2. Uniqueness in project 1 st & 2 nd year report text	28
1.4.3. Uniqueness in project final report text	28
2. Summary of textual analysis	29
2.1. Conclusions	29
2.2. Application	31
3. Comparison of D1.4 to D2.2 and D2.3	34
4. Conclusions and recommendations	34
5. References	37

Executive Summary

This report D4.1 “Report on synthesized results for the development of future DUT programme” presents synthesized results for the development of future DUT programme based on findings in WP4 Strategic support for DUT, WP1 ENUAC Cross-research Community deliverable report D1.4. “Final showcase of regional, national, European, and international projects, ideas, initiatives on the topic of accessibility and connectivity – the ENUAC projects”, WP2 Research Synthesis deliverable report D2.2. Analytical framework and methodology” and deliverable report D2.3 “Research synthesis”. All deliverables report D4.1, D1.4 and D2.2 & D2.3 examine 15 funded ENUAC (ERA-NET Urban Accessibility and Connectivity) projects focused on accessibility and connectivity.

The findings of this report D4.1 “*Report on synthesized results for the development of future DUT programme*” cover text analytics, including word frequency analysis, similarity analysis, word co-occurrence analysis, unique terms identification and text gap analysis of 15 ENUAC projects proposals, progress and final reports in comparison to the “Driving Urban Transition (DUT) towards a Sustainable and Liveable Urban Future” and “JPI Urban Europe’s Strategic Research and Innovation Agenda (SRIA)” texts.

Based on finding in WP4, the text analytics illustrates how these techniques enhance understanding of the project's progression, evolving focus, and alignment with DUT and SRIA objectives. Text analysis techniques improve the efficiency of processing and interpreting large volumes of project documentation. Research analysis indicates that it enables a more thorough interpretation of the project's shifting focus, aids in uncovering potential biases within the documents, and offers a clearer view of the connections and relationships across various project stages.

Text-analytic methods provide a powerful framework for both pre-funding and post-funding evaluation stages in project analysis. Prior to any funding decision, these methods enable a rapid, data-driven analysis of project proposals against the DUT and SRIA. This pre-funding stage allows evaluators to efficiently assess proposals for compliance with strategic goals, identify thematic trends (e.g., research topics, methodologies, geographic focus), and detect potential biases or gaps in the proposed research areas. Using these techniques before funding allocation ensures that selected projects closely align with DUT and SRIA objectives.

In the post-funding phase, text-analytic methods continue to be valuable for evaluating midterm and final reports. By comparing these reports to initial project goals and the DUT/SRIA frameworks, evaluators can check if the project objectives are being met and if any new research gaps or emerging themes have developed over time. This ongoing analysis supports strategic alignment and adaptation as projects progress. The WP4 results outline the main thematic clusters for DUT, SRIA, and ENUAC projects according to their year of implementation, along with their compliance, uniqueness, and gaps among the projects. These insights offer a clear view of project alignment and innovation both before funding decisions are made and throughout the project lifecycle.

Based on finding in WP1 and WP2 the analysis of the survey responses on the practical application of project knowledge in a variety of themes was done and the potential of the ENUAC portfolio was evaluated through two main aspects. D1.4 adopts a hands-on perspective, concentrating on the practical execution of these projects, the challenges encountered, and immediate research needs. In contrast, D2.2 & D2.3 take a more strategic view, focusing on overall project structures and long-term direction. D1.4 offers immediate, project-level insights, while D2.2 & D2.3 provide recommendations for structural adjustments and future guidance.

Together, D4.1, D1.4 and D2.2 & D2.3 deliver complementary perspectives, combining practical advice with strategic oversight to support effective, goal-aligned planning for future DUT projects.

Introduction

JPI Urban Europe's EN-UAC **Knowledge Hub ACUTE** (*Accessibility and Connectivity knowledge hub for Urban Transformation in Europe*) is a platform that enables exchange of knowledge, expertise and results in urban Accessibility and Connectivity from various actors [1]. ACUTE WP4 *Strategic Support for DUT* is designed to make knowledge collected in ENUAC projects & beyond available as support for all stakeholders in the future Horizon Europe Partnership Program, Driving Urban Transition (DUT) towards a Sustainable and Liveable Urban Future [2].

Objectives of the ACUTE WP4 are:

- Identify further needs for action to advance the topic of Urban Accessibility and Connectivity and embed it well in the future DUT Partnership Program.
- Synthesize results to provide input for the development of the future DUT programme in a way that fills research and implementation gaps and provide inputs to set agendas and create roadmaps.
- Transfer identified needs on call design and applicant support.
- Contribute to a holistic and wider system understanding of research results impact that goes beyond singular projects or activities.

The following **tasks** of ACUTE WP4 are defined:

- Task 4.1. Iterative identification of needs of programme bodies and specifications of topics required for. Task co-lead: LBTU (LLU), LU, participation of all partners.
- Task 4.2. Workshop/living lab with programme bodies, practitioners and experts to meet funders' needs, which are conducted in WP1-WP3. Task co-lead: LBTU (LLU), LU, participation of all partners.

Topics of experts' interviews and workshops, and inputs from WP1-WP3:

- Research gaps (elaborated in activities outlined above).
- Implementation gaps (perceived from researchers, intermediaries and public domain).
- Existing results to build on in further calls.
- New developments and trends in EN-UA-innovations and social practices, etc.
- Synergies with other funding programmes, initiatives specific on critical urban sectors of DUT: The 15-Minute City Transition Pathway (15minC), The Circular Urban Regenerative Economies Transition Pathway (CUE), The Positive Energy Districts Transition Pathway (PED).
- Intervention mix to address all needs in the innovation systems resp. by (potential) partners and target groups.

Deliverables: D4.1. Provide report on synthesized results for the development of future DUT programme (from and with WP1-WP3). Report on synthesized results for the development of future DUT programme is based on WP1 – WP3 results:

- WP1 *ENUAC Cross-research Community* aims to support ENUAC in connecting the projects and initiatives of the ENUAC to align their results with the SRIA 2.0 and DUT goals: **D1.4. Final showcase of regional, national, European and international projects, ideas, initiatives on the topic of accessibility and connectivity** [3].
- The core of the WP2 *Research synthesis* is a qualitative, deductive and comparative analysis of the projects within the ENUAC portfolio. Identification of research and implementation gaps to enable

WP4 to provide the DUT with strategic advice: **D2.2 Analytical framework and methodology & D2.3 Research synthesis** [4, 5].

- The objective of WP3 *Practitioner interaction* is integration of practitioners' needs and corresponding preparation of relevant subject areas related to urban accessibility and connectivity: **D3.2. Framework for support material and activities with transformative potential on identified needs.**

Overview of analysis process is based on WP1 – WP4 results (see Figure 1 **Fehler! Verweisquelle konnte nicht gefunden werden.**).

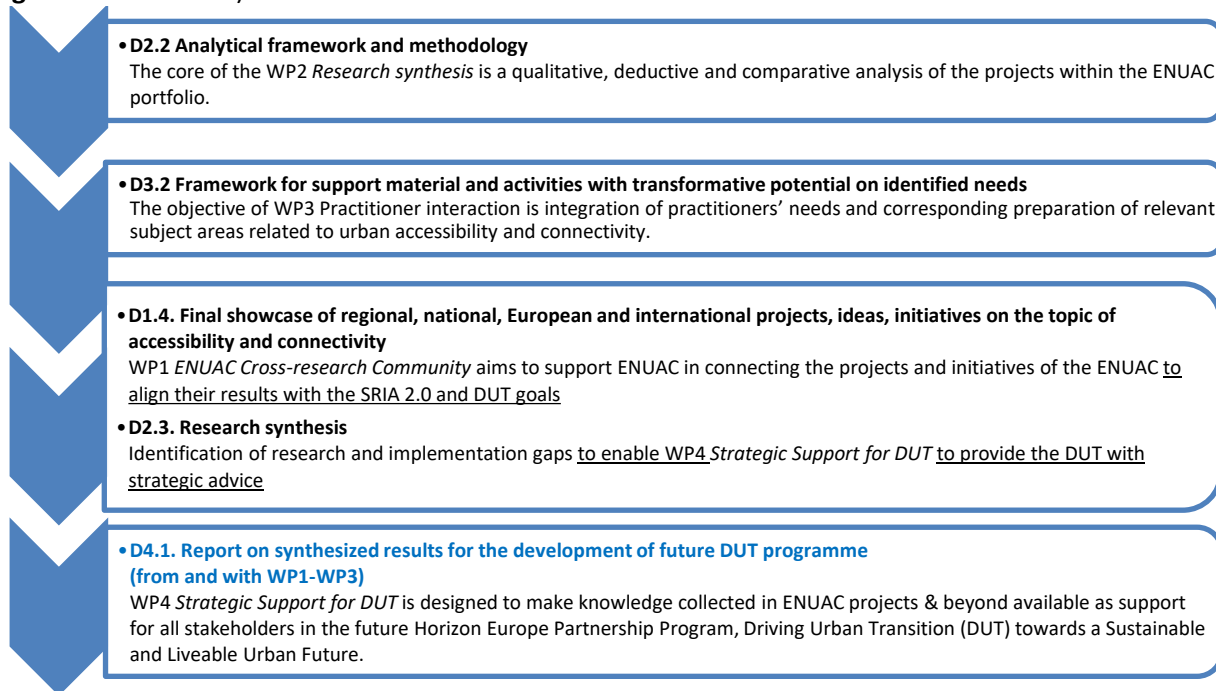


Figure 1: The overview of analysis process.

1. Textual analysis

Usually, the text is viewed as a means of conveying meaning, but it can also be seen as a structured sequence of words or an unstructured collection of words. These words can be represented in ways that enable analysis without being limited by grammatical structure. Texts can be subjected to statistical examination by exploring the interrelationships between words within a specific text and comparing these relationships to those observed in other texts. The approach with computational methods is [6]:

- Supplement to the qualitative analysis approach performed in WP1-WP3.
- Scalable – it can process tens and hundreds of documents (proposals, mid-term reports, final reports, etc.).
- Text analysis methods utilizes a “bag of words” technique, where the order of words does not matter.
- Data mining techniques to discover new knowledge for future calls, analysing word sequences in order.
- Manually adjustable, such as defining stop-words and precision.
- Not perfect, but it is semi-automated.

To develop a support methodology the following 15 ENUAC project' proposals and annual progress 1st & 2nd year reports were analysed:

- ASAP, *Awaken Sleeping Assets Project*: [ASAP \(smarturbanlogistics.eu\)](https://smarturbanlogistics.eu)
- CATAPULTS, *PoliCies for inclusive, demand-oriented and target group-specific automated mobility solutions for cities*: [CATAPULT – Policies for inclusive autonomous mobility solutions for cities \(catapultproject.eu\)](https://catapultproject.eu)
- COCOMO, *COmpeting and COmplementary MObility solutions in urban contexts*: [COmpeting and COmplementary MObility solutions in urban contexts \(COCOMO\) | Universiteit Utrecht | About CoCoMo \(uu.nl\)](https://uu.nl)
- DyMoN, *Dynamic Mobility Nudge: Shaping sustainable urban mobility behaviour with real-time, user-generated and public open data*: [Towards sustainable transport \(dymon.eu\)](https://dymon.eu)
- EASIER, *Seamless sustainable everyday urban mobility*: [EASIER \(dtu.dk\)](https://dtu.dk)
- EX-TRA, *EXperimenting with city streets to TRAnsform urban mobility*: [EX-TRA | Street Experiments \(ex-tra-project.eu\)](https://ex-tra-project.eu)
- GEOSENCE, *Geofencing strategies for implementation in urban traffic management and planning*: [GeoSence | Closer \(lindholmen.se\)](https://lindholmen.se)
- ITEM, *Inclusive Transition towards Electric Mobility*: [ITEM | Inclusive Transition to Electric Mobility \(ITEM\) \(itemresearch.org\)](https://itemresearch.org)
- JUSTICE, *Joining Urban morphology, Spatio-Temporal and socio-cognitive accessibility for an Inclusive City Environment*: [JUSTICE – Joining Urban morphology, Spatio-Temporal and socio-cognitive accessibility for an Inclusive City Environment \(justice-project.eu\)](https://justice-project.eu)
- MyFairShare, *Individual Mobility Budgets as a Foundation for Social and Ethical Carbon Reduction*: [MyFairShare | JPI Urban Europe](https://jpi-urban-europe.eu)
- SMARTHubs, *Smart Mobility Hubs as Game Changers in Transport*: [SmartHubs \(smartmobilityhubs.eu\)](https://smartmobilityhubs.eu)
- SORTEDMobility, *Self-Organized Rail Traffic for the Evolution of Decentralized MOBILITY*: [Sortedmobility](https://sortedmobility.eu)
- TAPforuncertainfuture, *Using Triple Access Planning to Enhance Urban Accessibility and Connectivity in the Face of Deep Uncertainty*: [Triple Access Planning for Uncertain Futures \(tapforuncertainty.eu\)](https://tapforuncertainty.eu)
- TuneOurBlock, *Transforming urban quarters to human scale environments: applying superblock concepts for different urban structure*: [TuneOurBlock | Smarter Than Car](https://smarterthanacar.eu)
- WALK Urban, *Walkable Urban Neighbourhoods – Freeing up Potential for Sustainable and Active Travel by Improving Walking and its Connections with Public Transport*: [Walk Urban](https://walkurban.eu)

The set of 15 ERA-NET Urban Accessibility and Connectivity (ENUAC) project' proposals, its 1st and 2nd year progress reports [7, 8], the “Driving Urban Transitions (DUT) to a Sustainable Future Roadmap 2022” [2] and “JPI Urban Europe’s Strategic Research and Innovation Agenda (SRIA 2.0.)” texts were analysed [9].

The evaluation and analysis included the following steps: gap, compliance, similarity and uniqueness (see Figure 2):

- **Gap** analysis is used to identify terms or topics that are missing or underrepresented in project documents relative to defined priorities:
 - *Text Gap Analysis*: identifies missing or underrepresented terms/topics across project documents in comparison to DUT and SRIA priorities. This is achieved through Term Frequency-Inverse Document Frequency (TF-IDF) analysis, highlighting terms prevalent in DUT/SRIA but absent or infrequent in project documents) [6, 10, 11].
 - *Word Co-occurrence Analysis*: Examines the frequency with which key terms from DUT/SRIA appear together in project documents, indicating gaps if certain key term pairings are missing.

- **Compliance** with DUT and SRIA by the following methods:
 - *Word Frequency Analysis*: compares the frequency of key terms/themes from DUT and SRIA across project documents, revealing how well project language aligns with strategic documents. Word clouds are generated to visualize the results.
 - *Text Similarity Analysis*: uses methods like Euclidean Distance or Hierarchical Clustering on Term-Document Matrices (TDM) to quantify alignment, grouping documents with similar thematic content, which suggests higher compliance with DUT and SRIA.
- **Similarity** is used to identify the closest documents, uncover specific groups or topics within a document set, and measure word or text similarity to cluster documents into distinct groups:
 - *Clustering*: clusters project documents based on shared language and themes, using methods like Hierarchical Clustering or K-means clustering on Document-Term Matrices (DTM). Projects grouped closely share greater thematic similarity. Text similarity analysis was done in two ways by clustering documents and words of DUT, SRIA and 15 ENUAC projects text. Clustering by documents helps identify the most closely related documents, while clustering by words reveals specific topics within a set of text documents.
 - *Dendrogram Visualization*: visualizes clusters and similarities across project documents, highlighting the closest relationships in content.
- **Uniqueness** is used to identify topics that are specific to each text:
 - *Unique Terms Identification (TF-IDF)*: TF-IDF values are used to highlight distinctive words in each project document, revealing unique themes or areas of focus that set each project apart from others) [6, 10, 11].
 - *Word Clouds and Frequency Analysis*: visual tools such as word clouds highlight frequent and unique terms within each project, allowing for quick identification of distinct thematic focus.



Figure 2: DUT, SRIA and 15 ENUAC projects text analytics scheme.

1.1. Pre-processing

For the documents text analysis, files are converted to the TXT format, and cleaning is an essential first step. Each project file (proposal or report) is in PDF format with a pre-defined structure. Since the structure is irrelevant to the analysis, the PDF files are converted to TXT format, and text cleaning procedures are applied (see Figure 3).

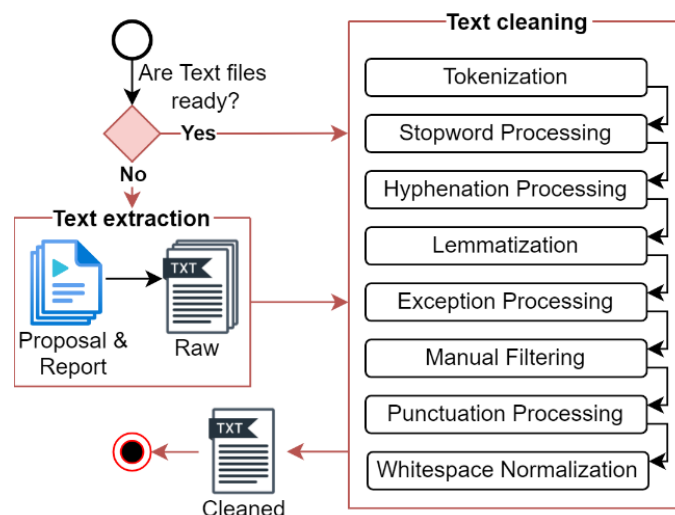


Figure 3: The text pre-processing scheme.

Text cleaning include following procedures:

- Read the content of the file.
- Tokenization: tokenize file into individual words.
- Remove stop words from the tokenized words.
- Split hyphenated words into separate tokens.
- Lemmatize each word based on its Part of Speech (POS) tag.
- Remove punctuation and numbers, except for terms in the exceptions list.
- Normalize whitespace by removing extra spaces.
- Keep words that are valid English words or in the exceptions list.
- Exclude words found in the “Manual filter” list.
- Save the cleaned text as a new file.

One of the libraries commonly used for text analysis in Python is the Natural Language Toolkit, or NLTK. This library provides tools and functions for processing and analysing text data. After files are converted to TXT format, the text is broken down into smaller units – words – In a process called tokenization. This step is basic but important for later tasks like classification or sentiment analysis. In NLTK, tokenization is done using the `nltk.tokenize.word_tokenize` function.

It's also important to remove stop words – common words like “and”, “the”, “is”, and “in” – that add little meaning to the text and can create unnecessary “noise.” NLTK has a predefined list of stop words in `nltk.corpus.stopwords`, which can be used to filter out these terms during tokenization. When working with hyphenated words like “well-known”, they can be split into separate words (“well” and “known”) if needed, especially for frequency analysis. In this case, hyphenated words are split into individual tokens using the `split` method.

To make sure different forms of a word (like “ran” and “running”) are recognized as the same word, lemmatization is applied using the `lemmatizer.lemmatize` function in NLTK. This process simplifies words to their base form, which reduces redundancy and makes the analysis more accurate. The correct Part of Speech (POS) tag for each word, which helps in accurate lemmatization, is identified using the `get_wordnet_pos` function along with `nltk.pos_tag`. Some words may not be stop words but might still be unnecessary for certain types of analysis. For example, common words in reports like “project” or “research” might not add much value in specific contexts, so they are removed manually. A similar approach is applied to specific terms that need to be kept, such as “PED”, “15mC”, or “CUE”, even though they may not be part of the default English corpus in NLTK.

1.2. Word frequency analysis

Word frequency analysis helps to compare documents by showing common themes, terms, and patterns in each text. By counting how often words appear, this analysis reveals main topics and areas of focus, making it easier to spot similarities and differences. For instance, if two documents are compared, word frequency analysis can show which terms are emphasized more in one text, hinting at a different focus or approach. Unique words in one document can signal specific themes or priorities that aren't in the other. Likewise, missing common terms in one document may reveal gaps or areas less explored compared to the other text.

This analysis also supports text similarity checks by identifying shared themes, helping to see if two documents cover similar topics or have related goals. It also highlights distinctive elements in each document, making comparisons more precise. Visual tools like word clouds and frequency plots make it easy to see prominent themes and terms, providing a clear, data-based look at each document's content. To create word clouds, R packages *wordcloud*, and *RColorBrewer* were used. The *wordcloud* package generates the visualizations, and *RColorBrewer* provides the color palettes for word cloud visualization, adding clarity and appeal. Analysis of the DUT, SRIA and ENUAC projects' report texts word frequency analysis is done by Visualizing Data through word cloud generation for "Driving Urban Transitions (DUT) to a Sustainable Future Roadmap 2022" and "JPI Urban Europe's Strategic Research and Innovation Agenda (SRIA 2.0)." texts (see Figure 4).

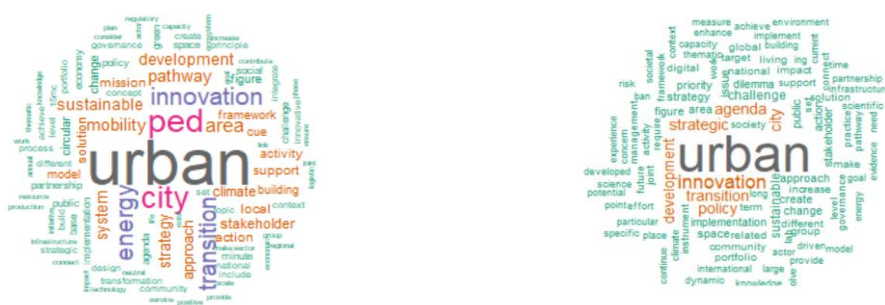


Figure 4: Word Clouds of DUT and SRIA texts.

The word clouds for DUT (Driving Urban Transitions) and SRIA (Strategic Research and Innovation Agenda) show both common themes and different areas of focus. Both emphasize terms like “urban”, “innovation”, “transition”, and “city”, indicating a shared commitment to sustainable cities.

However, each has a unique focus. The DUT word cloud includes words like “energy”, “sustainable”, “pathway”, “model”, and “mobility”, suggesting it is more focused on practical solutions in areas like energy and transportation. DUT aims to make real changes in cities through specific initiatives, such as Positive Energy Districts, the 15-Minute City, and Circular Urban Economies, which focus on cleaner energy, accessible neighbourhoods, and efficient resource use. On the other hand, SRIA highlights terms like “policy”, “agenda”, “challenge”, and “implementation”, showing a focus on creating research and policy strategies to guide urban development. SRIA’s role is to set priorities, encourage collaboration, and help address complex urban issues that may have conflicting goals, aiming to find balanced solutions.

In summary, DUT focuses on practical actions for immediate impact in cities, while SRIA provides strategic guidance and research priorities to support these changes at a broader policy level. Together, they work toward the common goal of sustainable urban development.

1.2.1. Analysis of project proposal's report texts

Word frequency analysis was done by visualizing data through word cloud generation for 15 ENUAC projects proposals (see Figure 5).



ASAP



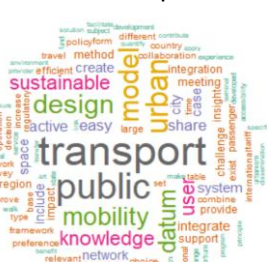
Catapult



COCOMO



DvMoN



Easier



Ex-TRA



Geo-Sence



ITEM



JUSTICE



MyFairShare



SmartHubs



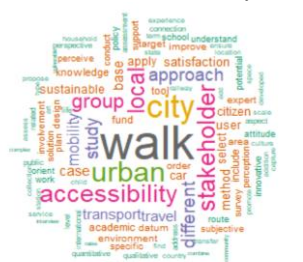
SortedMobility



TAP



TuneOurBlock



WalkUrban

Figure 5: Word Clouds of 15 ENUAC projects proposals report texts.

The word clouds for the 15 ENUAC project proposals show both shared themes and unique focuses for each project. Most word clouds highlight common words like “mobility”, “urban”, “city”, “transport”, and “sustainable.” This shows that the projects are closely related, with a common goal of improving urban mobility, promoting sustainable transport, and solving city-specific challenges.

Each project’s word cloud also includes unique words that point to specific areas of focus within the larger theme of urban mobility:

- ASAP mentions “logistics” and “urban solutions”, suggesting it focuses on city logistics.
- Catapult and ITEM emphasize “hub” and “mobility”, possibly focusing on centralizing transport services.
- SortedMobility includes “railway” and “algorithm”, indicating a focus on rail systems and data-driven traffic management.
- Ex-TRA highlights “accessibility” and “street”, suggesting a focus on making streets easier to access.
- MyFairShare shows “budget” and “mobility”, which might mean it focuses on the budget of mobility.

The projects are closely related, with all focusing on urban mobility and sustainability. However, each one tackles a unique aspect, such as logistics, budgeting, accessibility, or data solutions. This variety shows that, while they share similar goals, each project has a specific angle or area of expertise.

1.2.2. Analysis of the project’s 1st & 2nd year report texts

Word frequency analysis was done by visualizing data through word cloud generation for the 15 ENUAC project’s 1st & 2nd year reports (see Figure 6).

Comparing the original project proposals with the word clouds from the 1st & 2nd year reports reveals some clear shifts in focus as the projects developed. In the 1st and 2nd year reports, words related to implementation and real-life application, like “test”, “pilot”, “policy”, “stakeholder”, and “impact”, appear more often. This suggests that the projects have shifted from planning and concepts to practical, actionable steps, moving closer to real-world applications and testing.

Words like “group”, “stakeholder”, “consortium”, and “participation” are more common, indicating a stronger focus on collaboration and involving various stakeholders. This likely reflects a need for more engagement with people and organizations as the projects progress into active phases that require community input or expert support.

The presence of terms like “algorithm”, “data”, “platform”, and “simulation” shows a greater use of tech and data tools to tackle urban mobility issues. Additionally, words like “policy”, “regulation”, and “governance” appear more frequently, maybe indicating that projects are now more aligned with local and national policies.



ASAP



Catapult



COCOMO



DyMoN



Easier



Ex-TRA



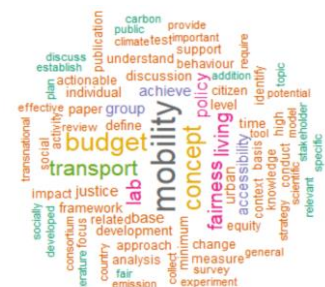
Geo-Sence



ITEM



JUSTICE



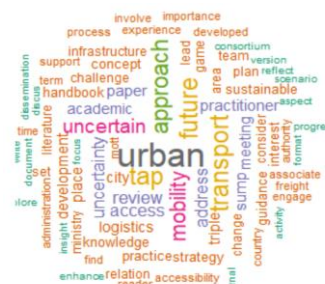
MyFairShare



SmartHubs



SortedMobility



TAP



TuneOurBlock



WalkUrban

Figure 6: Word Clouds of 15 ENUAC projects 1st & 2nd year report texts.

Some projects show a shift toward specific issues or infrastructure areas, likely informed by early findings and challenges faced during implementation. For example:

- SortedMobility focuses on “railway” and “traffic”, showing a focus on transport systems.
- Ex-TRA emphasizes “accessibility” and “street”, suggesting a focus on making public spaces easier to access.
- DyMoN mentions “parking” and “traffic”, pointing to a targeted approach to managing urban mobility issues.

Some projects now highlight specific terms like “school”, “target”, or “local”, indicating they are focusing on particular community needs. For example:

- WalkUrban focuses on “school”, likely to improve safe routes for students.
- TuneOurBlock shows more local engagement terms, suggesting a neighbourhood-level focus.

From proposals to reports, projects have shifted from broad ideas to hands-on actions. They're now more focused on real-world applications, working with communities, using tech and data, addressing specific issues, and aligning with policy needs. This evolution shows each project adapting to real-world demands as they move forward.

1.2.3. Analysis of project final report texts

Word frequency analysis was done by visualizing data through word cloud generation for 8 ENUAC projects final reports (see Figure 7) as of September 2024 only 8 final project reports were available for the analysis: ASAP, Catapult, COCOMO, DyMoN, Ex-TRA, Geo-Sence, ITEM, JUSTICE, SmartHubs, TAP, WalkUrban.



Figure 7: WordClouds of 8 ENUAC projects final report texts.

The word clouds for the final reports reflect shifts in focus as the projects near completion. For some of them there is a clear emphasis on finalizing results and producing resources, while others maintain a strong focus on practical applications and community engagement.

In the final reports, terms like “report”, “publish”, “document”, “guidance”, “recommendation”, “outcome”, and “deliverable” appear more often, indicating a focus on consolidating findings, sharing results, and producing final documentation. For instance:

- SmartHubs includes terms like “deliverable”, “integration”, and “publish” showing a focus on finalizing results and sharing findings.
- TAP highlights “handbook” and “guidance”, suggesting the creation of practical resources or final reports.

Some projects maintain a strong emphasis on engagement and practical applications, with terms like “meeting”, “workshop”, “session”, and “community” frequently appearing. This suggests ongoing collaboration with stakeholders and a focus on real-world testing:

- WalkUrban includes terms like “conference”, “school”, and “session”, indicating a final focus on community engagement, possibly with educational elements.
- Ex-TRA continues to emphasize “accessibility” and “experiment”, showing ongoing attention to urban accessibility solutions.

There is a notable trend toward producing resources that can be shared and used by others. Words like “publish”, “guidance”, “tool”, and “report” highlight the effort to share knowledge widely:

- ASAP and JUSTICE use terms like “tool” and “data”, focusing on providing practical tools and data resources.
- COCOMO and TAP emphasize “handbook” and “guidance”, aiming to provide best practices and comprehensive guides based on their findings.

Compared to the initial proposals and early reports, the final reports have a greater emphasis on producing usable outcomes like guides, handbooks, and recommendations. Terms such as “guide”, “final”, “deliverable”, and “handbook” are more prominent, reflecting a shift from early-stage concepts and testing to providing completed resources for wider application. The final reports are focused on wrapping up each project, producing resources, and sharing practical outputs like guides, handbooks, and recommendations. While stakeholder engagement and practical testing remain key for some projects, others are more centred on consolidating knowledge for broader use. This final stage marks a progression from initial planning and testing to creating resources that others can adopt, ensuring the projects’ findings have lasting impact.

1.3. Similarity analysis

Text similarity analysis measures how closely related words or documents are, grouping them into clusters of similar items. The goal is to find groups of documents that are very similar to each other, making it easier to identify patterns and themes within each group. This analysis can be used to cluster documents based on shared topics, but it can also cluster words, showing connections between different terms.

The analysis starts by calculating the Euclidean Distances [12] between documents, based on how often specific terms appear in each of it. A lower Euclidean distance indicates that the two documents have more similar content. A higher Euclidean distance indicates that the documents are more dissimilar. Then hierarchical clustering groups documents that have a similar word usage. Hierarchical Clustering builds a tree of clusters, where each document starts in its own cluster and is progressively merged with others. The results are displayed as a dendrogram, a tree-like diagram that shows how documents or words are related

within clusters. This process is applied to both a Term-Document Matrix (TDM) and a Document-Term Matrix (DTM).

The TDM focuses on relationships between words across documents, helping to identify related keywords or common themes across different projects. The DTM on the other hand focuses on how each document uses different words, which helps in comparing the documents themselves. Using both matrices allows for a deeper comparison of both terms and documents. Text similarity analysis follows these steps:

- Load necessary libraries and the text data.
- Create a Term-Document Matrix (TDM).
- Convert the TDM into a data frame.
- Standardize the data for consistency.
- Calculate the Euclidean Distances (ED) between documents.
- Use hierarchical clustering (Ward's method) to group similar documents [12].
- Visualize clusters with a dendrogram.
- Create a Document-Term Matrix (DTM).
- Repeat steps 3 to 7 for the DTM.

Text similarity analysis helps to compare project documents by grouping those with similar language, themes, or topics. This makes it easier to spot projects with similar goals, methods, or results, highlighting areas where they could work together. For example, projects focused on “mobility” and “sustainability” might be grouped together, showing shared goals. The analysis can also identify unique groups, showing which projects stand out with different perspectives or innovative ideas. **This helps to spot projects that bring something new to the portfolio.** By examining words across projects, text similarity analysis reveals which topics are widely covered and which need more attention, helping to balance the project portfolio. In short, this analysis organizes and compares documents, giving insights into common themes, unique ideas, and possible gaps.

1.3.1. Relationships of project proposal's report text

The clustering results by project proposals text is shown by the following dendrogram in Figure 8. Based on the analysis, it can be concluded that the proposal text similarity from low to high is as follows: the projects with the highest similarity to the text of their reports are as follows: CATAPULT and SmartHub, then Ex-TRA, DyMoN, COCOMO, MyFairShare, GeoSense, ASAP, EASIER, WalkUrban, TuneOurBlock, TAP, SortedMobility, JUSTICE and ITEM.

It's important to note that the dendrogram shows textual similarity rather than thematic similarity. Projects are grouped based on the language and terms used in their proposal texts, not necessarily because they have the same core focus.

To derive meaningful insights from cluster analysis, word clouds serve as an essential preliminary tool that enables visualizing the most frequently occurring terms in each project grouping. Cluster analysis alone provides an initial organization of projects based on textual similarity, yet understanding the nuanced connections between these projects requires the integrated use of TF-IDF, word frequency analysis, and traditional methods like abstract reviews.

For example, word clouds give an accessible overview of dominant terms in each cluster, offering an initial sense of thematic trends. However, to understand why certain projects align within a cluster, TF-IDF analysis is needed to pinpoint unique terms and thematic priorities that define each grouping in greater detail. Word frequency analysis further supplements this by highlighting the most common themes across clusters, clarifying overarching topics and shared focuses.

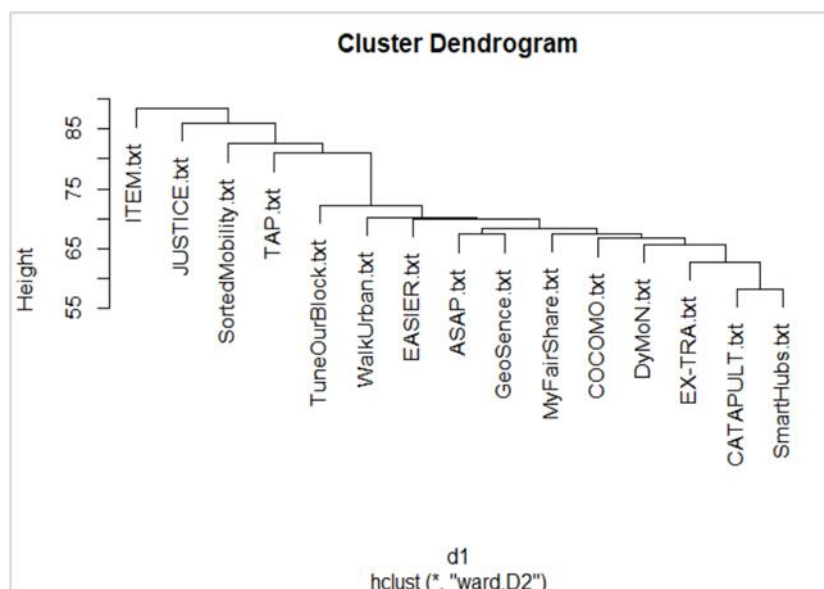


Figure 8: Dendrogram of 15 ENUAC projects proposal texts (Height - Euclidean Distances (ED) between clusters).

Without this combined approach, each method on its own offers a partial view. Cluster analysis groups projects, but without TF-IDF, it cannot indicate which terms are most distinctive or relevant within each cluster. Similarly, word clouds and frequency distributions provide prominent term snapshots, yet they do not reveal structural relationships or finer thematic distinctions that clustering can provide.

Challenges for cluster analysis are interpretability of the clusters or understanding the meaning of a cluster or the reasons behind why certain documents are grouped together. Projects with more similar text are in one cluster, for example CATAPULT and SmartHubs. The generated TF-IDF values help us to see the most prominent terms for the CATAPULT - SmartHubs cluster (see Figure 9. In both documents words like “lab”, “living”, “station”, “cargo”, “logistic”, “game”, “shuttle” and “traveller” with a high IT-IDF value occur in the project proposal text. CATAPULT - SmartHubs cluster potentially centred to shared resources, logistic in the city or centralized solutions for urban mobility.

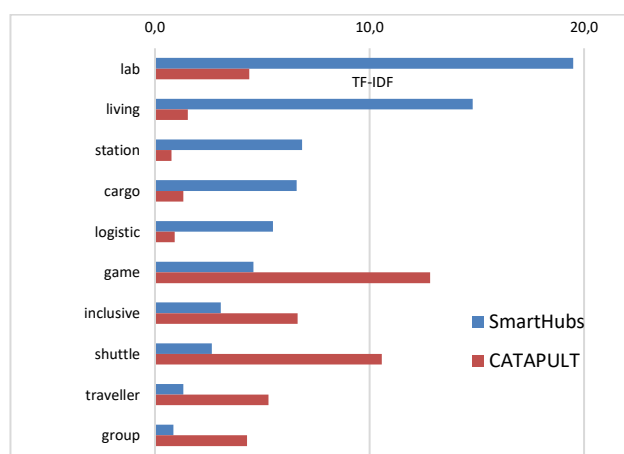


Figure 9: TF-IDF unique words for CATAPULT - SmartHubs cluster, based on 15 ENUAC projects proposal analysis.

For the analysis of the project topics the cluster analysis must be supplemented with TF-IDF, word frequency analysis and traditional methods (read the abstract of the document). Using a similar approach, the following clusters were identified – groups of projects that are closely related based on the similarity of their textual content:

- Cluster 1: ITEM, JUSTICE, SortedMobility, TAP projects are closely related, possibly sharing themes around infrastructure, public transport, and system-level strategies, as suggested by their tight grouping. By comparing the word clouds of each project, the recurring terms that appear prominently across the individual clouds can be analysed. Frequent words like “mobility”, “policy”, or “transport” shared across these projects indicate shared thematic focuses, supporting an understanding of the cluster’s themes. If all four projects have words like “public”, “transport”, and “system” frequently, it can be suggested they might share a focus on public transport systems.
- Cluster 2: TuneOurBlock, WalkUrban, EASIER remain closely linked, likely indicating a shared focus on community engagement, neighborhood accessibility, or local-level interventions within urban areas.
- Cluster 3: ASAP, Geo-Sence, MyFairShare, COCOMO, DyMoN projects may focus on broader urban challenges, possibly covering data-driven solutions, environmental sustainability, and various urban mobility approaches. While their goals may vary, they share a common ground in addressing complex urban issues with diverse strategies.
- Cluster 4: EX-TRA, CATAPULT, SmartHubs form a cohesive group, potentially centred around hub-based infrastructure or shared mobility services. These projects likely have a strong emphasis on transport hubs, shared resources, or centralized solutions for urban mobility.

The clustering result by project proposals words is shown by the following dendrogram (see Figure 10):

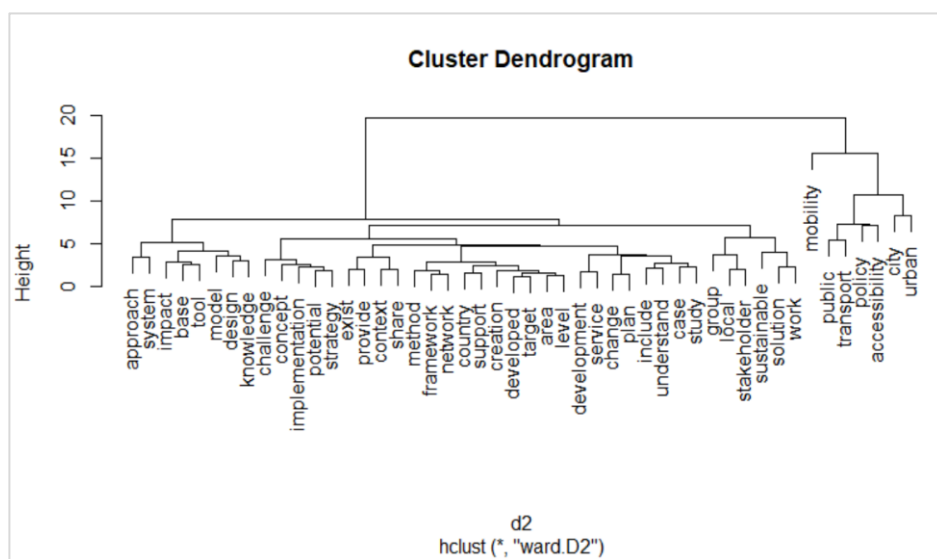


Figure 10: Dendrogram of 15 ENUAC projects proposals words (Height - ED between clusters).

Based on the dendrogram, it can be inferred that the most similar words can be organized into the following clusters of similarity:

- Cluster 1: “approach, system, impact, base, tool, model, design, knowledge”. This cluster suggests a focus on methodologies and foundational tools, indicating that many projects discuss systematic approaches and the knowledge base they build upon.
- Cluster 2: “challenge, concept, implementation, potential, strategy”. This group reflects discussions on overcoming barriers and strategic planning, possibly related to how projects approach urban mobility issues.
- Cluster 3: “exist, provide, context, share, method, framework, network, country, support, creation, developed, target, area, level, development, service, change, plan, include, understand, case, study”. This is a larger cluster focused on structural and contextual terms, suggesting a broad theme of how

projects provide support frameworks, target specific areas, and implement changes. It may relate to the operational aspect of project implementation and collaboration.

- Cluster 4: “group, local, stakeholder, sustainable, solution, work”. This cluster highlights a focus on local stakeholder involvement and sustainable solutions, indicating the community-focused nature of many projects.
- Cluster 5: “public, transport, policy, accessibility”. This group likely represents discussions around public accessibility, transport policy, and inclusivity, core themes for urban mobility projects.
- Cluster 6: “city, urban”. This small cluster shows the core setting for all projects, confirming their urban focus.
- Cluster 7: “mobility”. This standalone cluster reflects the central theme of all projects, emphasizing that mobility is the key focus across the board.

The clustering results by project proposals text, DUT and SRIA texts is shown by the dendrogram in Figure 11:

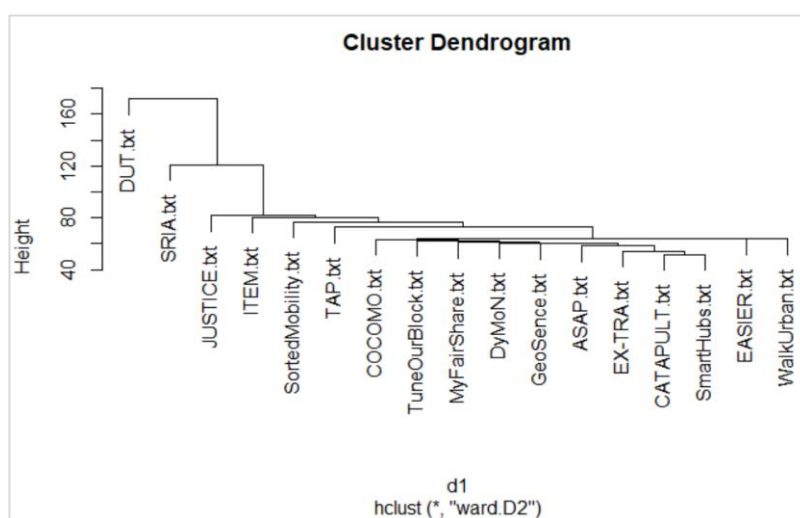


Figure 11: Dendrogram of 15 ENUAC projects proposal texts, DUT and SRIA texts (Height – ED between clusters).

DUT and SRIA texts cluster closely with each other but remain somewhat distinct from the individual ENUAC project proposals. This separation suggests that, while DUT and SRIA share some language and thematic content, they have a broader or more strategic focus compared to the project-specific proposals. The close pairing between DUT and SRIA also indicates that these documents are more aligned with each other than with individual project proposals, likely because they set overarching goals, principles, and frameworks for urban transitions.

Based on the comparison with DUT and SRIA, it can be concluded that the projects exhibiting the highest textual similarity are as follows: CATAPULT and SmartHub, then Ex-TRA, ASAP, DyMoN, GeoSence, MyFairShare, TuneOurBlock, COCOMO, EASIER, WalkUrban, TAP, SortedMobility, ITEM, JUSTICE, SRIA and DUT.

JUSTICE and ITEM form the closest link to the DUT and SRIA cluster, suggesting that these projects might align more with the strategic and policy-oriented language or goals set out in the DUT and SRIA documents. They may include higher-level themes related to governance, policy impact, or systemic change, which resonates with the broader focus of DUT and SRIA. The other projects are more closely grouped together but remain distinct from DUT and SRIA.

These clusters likely reflect a more practical, hands-on approach found in project proposals as opposed to the strategic language seen in DUT and SRIA. Projects like SortedMobility, TAP, COCOMO, and TuneOurBlock

form a cohesive sub-group, indicating shared themes or vocabulary that differ from DUT and SRIA's higher-level approach.

The K-means clustering result by project proposals, DUT and SRIA words is shown by the next dendrogram (see Figure 12).

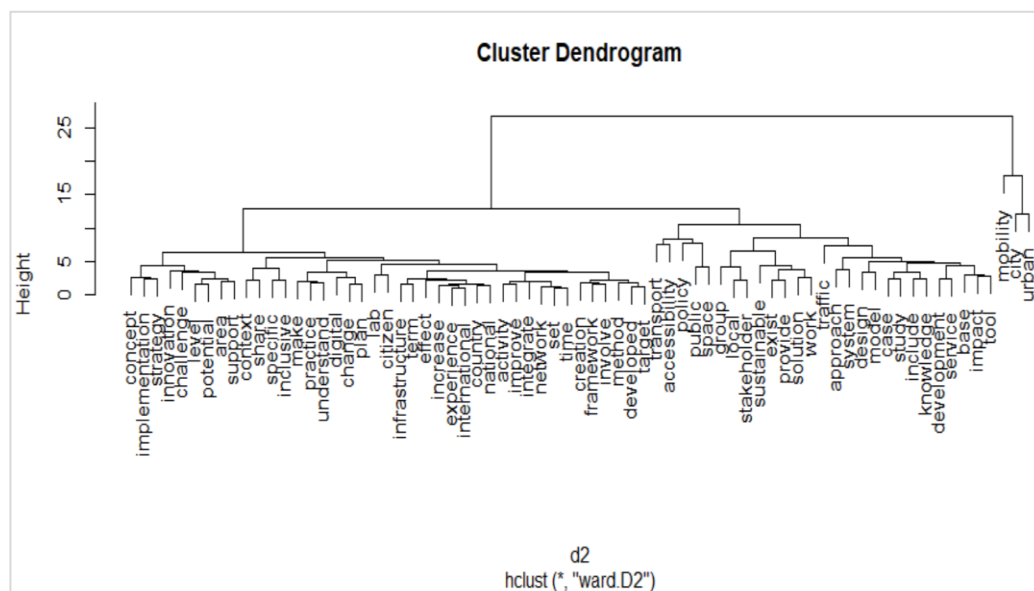


Figure 12: Dendrogram of 15 ENUAC projects proposals, DUT and SRIA words (Height - ED between clusters).

In the word dendrogram, high-level terms like **“mobility”**, **“urban”**, **“policy”**, and **“sustainable”** are clustered, reflecting shared central themes across DUT, SRIA, and the project proposals. These terms are foundational to both the individual projects and the strategic objectives set by DUT and SRIA. Terms such as **“implement”**, **“concept”**, **“change”**, **“infrastructure”**, and **“framework”** are grouped, indicating their importance across all documents but especially in DUT and SRIA. These words may represent themes that are more abstract and policy-focused, aligning with the overarching guidance provided by DUT and SRIA.

Clusters with words like **“local”**, **“stakeholder”**, **“solution”**, and **“public transport”** show a focus on practical, community-oriented aspects, which are more likely emphasized in project proposals than in strategic documents like DUT and SRIA. These clusters represent the hands-on elements specific to individual projects. **“Mobility”** and **“urban”** appear as a distinct cluster, underscoring their central importance across all documents. This reflects the core focus of the ENUAC projects, DUT, and SRIA on transforming urban mobility systems. It can be inferred that words sharing the greatest similarity can be grouped together into the following clusters of similarity:

1. Cluster: concept, implementation, strategy, innovation, challenge, level, potential, area, support, context, share, specific, inclusive, make, practice, understand, digital, change, plan, citizen, infrastructure, term, effect, increase, experience, international, country, national, activity, improve, integrate, network, set, time, creation, framework, involve, method, developed, target;
2. Cluster: transport, accessibility, policy, public, space, group, local, stakeholder, exist, provide, solution, work, traffic, approach, system, design, model, case, study, include, knowledge, development, service, base, impact, tool;
3. Cluster: mobility, city, tool.

1.3.2. Relationships of project 1st & 2nd year report text

The clustering results by project's 1st & 2nd year report texts is shown by the dendrogram in Figure 13. Based on the analysis, it can be concluded that the projects with the highest text similarity are as follows: CATAPULT

and WalkUrban, then Ex-TRA, JUSTICE, SmartHub, GeoSense, TuneOurBlock, COCOMO, ITEM, then DyMoN, ASAP, TAP, EASIER, MyFairShare, and SortedMobility.

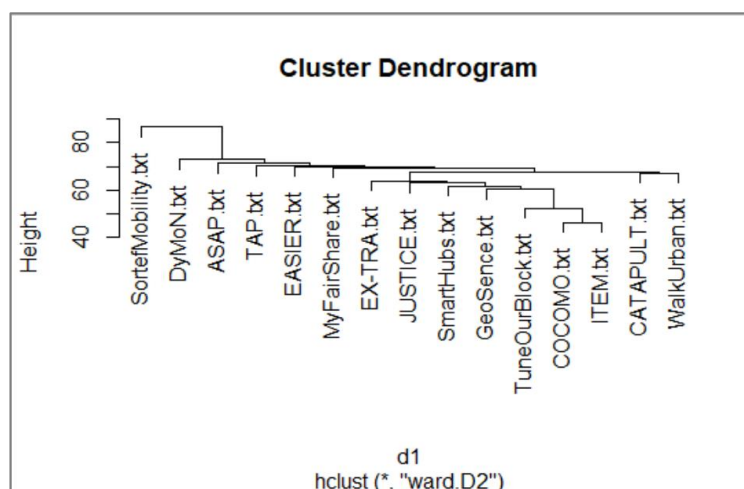


Figure 13: Dendrogram of 15 ENUAC projects 1st & 2nd year report texts (Height - ED between clusters).

SortedMobility and DyMoN are now more closely grouped, showing a similarity that may have emerged during the project progress. This could suggest that both projects have adopted similar terms or focus areas, possibly due to similar challenges or overlapping approaches in their implementations. ASAP, TAP, and EASIER are now in closer proximity.

This might indicate that these projects have converged in their focus or language during the first two years, possibly aligning around certain operational or practical themes. CATAPULT and ITEM form a distinct cluster near WalkUrban. This grouping may reflect a shared focus on infrastructure or mobility solutions that became more pronounced during the project execution phases.

The clustering result by project's 1st & 2nd year report words is shown by dendrogram in Figure 14.

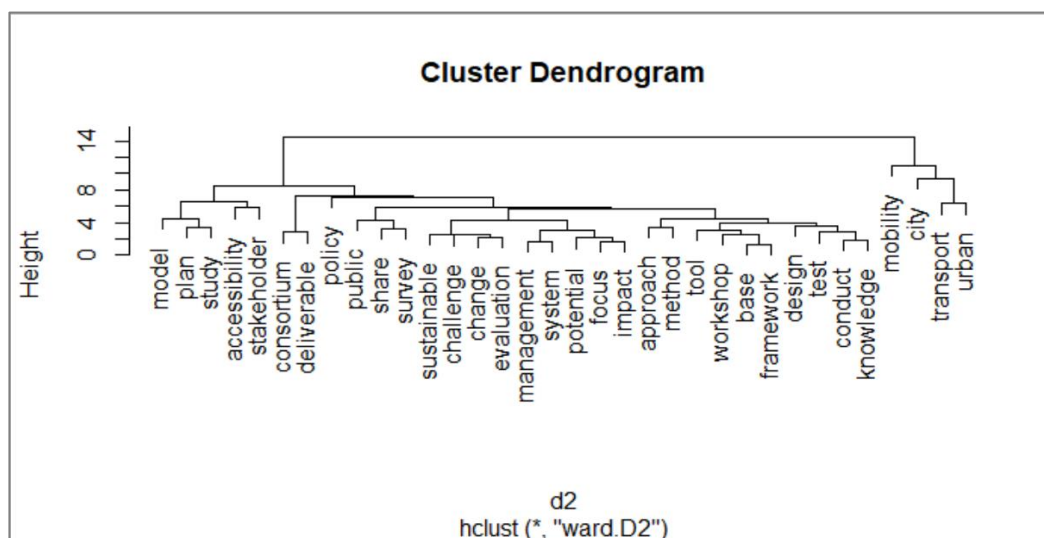


Figure 14: Dendrogram of 15 ENUAC projects 1st & 2nd year report words (Height - ED between clusters).

Terms such as **"stakeholder"**, **"policy"**, **"evaluation"**, and **"sustainable"** are now more prominent in clusters. This suggests a growing emphasis on stakeholder involvement, policy alignment, and sustainable practices as projects moved into their later stages. These terms reflect themes that would naturally become more important during real-world implementation and evaluation. Words like **"workshop"**, **"school"**,

“framework”, “approach”, and “model” are now closely grouped, indicating a focus on practical applications, workshops, and structured frameworks for testing and evaluating outcomes.

This cluster highlights the importance of structured interactions, such as workshops or stakeholder meetings, as key components of the projects in their 1st & 2nd years. Core words like “mobility”, “urban”, and “transport” remain clustered and central, showing that these fundamental themes continue to be the backbone of all project discussions. However, they are now closely associated with terms like “public”, “policy”, and “accessibility”, indicating that mobility is increasingly seen in relation to public access and policy considerations.

Based on the dendrogram, it can be inferred that the most similar words can be organized into the following clusters of similarity:

1. Cluster: model, plan, study, accessibility, stakeholder;
2. Cluster: consortium, deliverable;
3. Cluster: policy, public, share, survey, sustainable, challenge, change, evaluation, management, system, potential, focus, impact;
4. Cluster: approach, method, tool, workshop, base, framework, design, test, conduct, knowledge;
5. Cluster: mobility, city, transport, urban.

The clustering results by project 1st & 2nd year report text, DUT and SRIA texts is shown by dendrogram (see Figure 15).

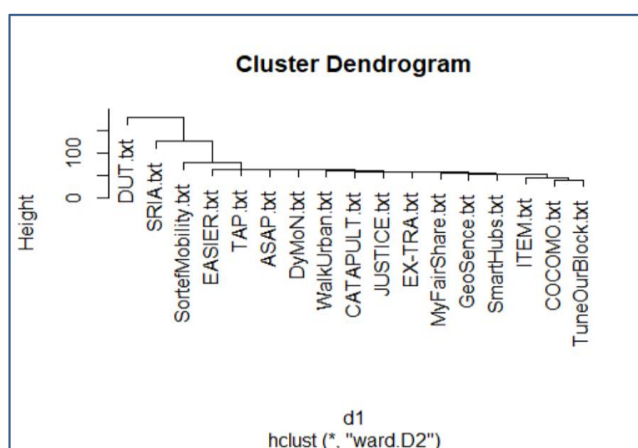


Figure 15: Dendrogram of 15 ENUAC projects 1st & 2nd year report texts, DUT and SRIA texts (Height - ED between clusters).

DUT and SRIA cluster closely together, just as they did in the proposal phase analysis. This tight grouping confirms that both documents share similar high-level language and strategic themes, which continue to be distinct from the majority of the ENUAC project reports. The strategic and overarching language used in DUT and SRIA is set apart from the more practical, specific focus found in the 1st & 2nd year project reports. This indicates that, even as projects moved closer to real-world implementation, their language and focus did not fully converge with the broader frameworks provided by DUT and SRIA.

Based on the comparison with DUT and SRIA, it can be concluded that the projects exhibiting the highest similarity are as follows: ITEM, COCOMO, TuneOurBlock, then EASIER, TAP, ASAP, DyMoN, WalkUrban, CATAPULT, JUSTICE, Ex-TRA, MyFairShare, GeoSence, SmartHub, then SortedMobility, SRIA and DUT.

The clustering result by project 1st & 2nd year report, DUT and SRIA words is shown by dendrogram (see Figure 16).

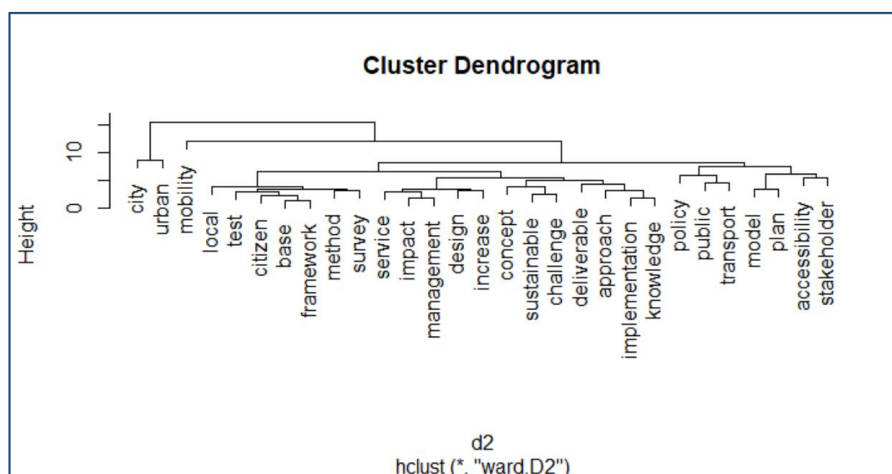


Figure 16: Dendrogram of 15 ENUAC projects 1st & 2nd year report, DUT and SRIA words (Height - ED between clusters).

The words “city”, “urban”, and “mobility” were clustered together, so showing these fundamental themes continues to be the backbone of all document discussions. Terms such as “stakeholder”, “policy”, “public”, “accessibility”, and “transport” are included in one cluster and are now more prominent.

Words such as “implementation”, “concept”, “framework”, “methods”, and “survey” are now grouped into one group, and it is indicated a practical applications, workshops, and structured frameworks to test and evaluate results. It can be inferred that words sharing the greatest similarity can be grouped together into the following clusters of similarity:

1. Cluster: city, urban, mobility;
2. Cluster: implementation, concept, framework, methods, and survey and all other words;
3. Cluster: public, transport and accessibility.

1.3.3. Relationships of project’s final report texts

The clustering results by 8 projects final report text is shown by the dendrogram in Figure 17. Based on the analysis, it can be concluded that the projects with the highest text similarity are as follows: CATAPULT, WalkUrban, Ex-TRA and SmartHub, then ASAP & TAP, and COCOMO & DyMoN.

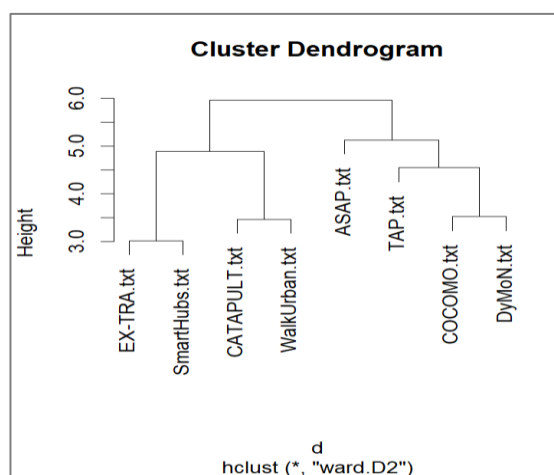


Figure 17: Dendrogram of 8 ENUAC projects final report texts (Height - ED between clusters).

The clustering result by 8 ENUAC project final report words is shown by the dendrogram (see Figure 18).

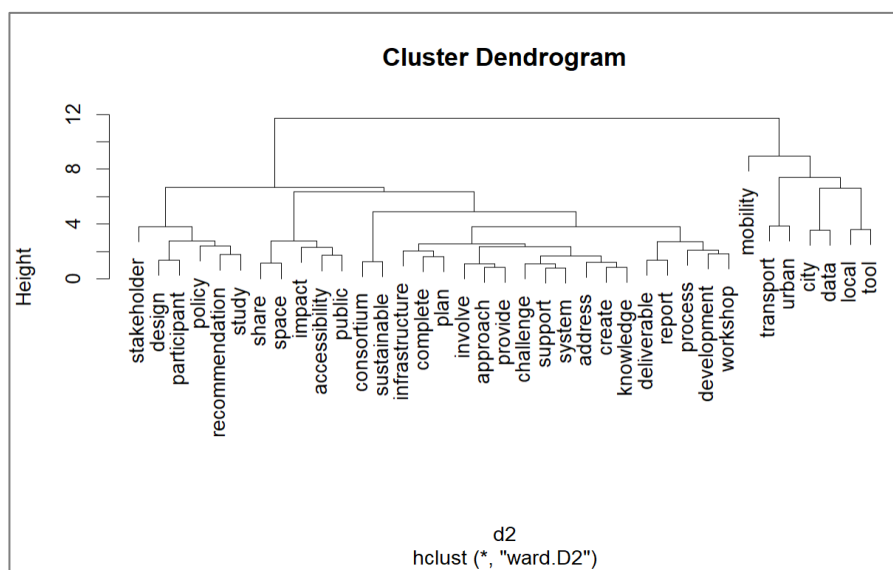


Figure 18: Dendrogram 8 ENUAC projects final report words (Height - ED between clusters).

Based on the dendrogram, it can be inferred that the most similar words can be organized into the following clusters of similarity:

1. Cluster: Recommendation & infrastructure;
2. Cluster: Data, tool & mobility.

The clustering results by 8 project final report text, DUT and SRIA texts is shown by dendrogram (see Figure 19). Based on the comparison with DUT and SRIA, it can be concluded that the projects exhibiting the highest similarity are as follows: DyMoN, CATAPULT, SmartHub, then ASAP, WalkUrban then Ex-TRA, COCOMO and SRIA with DUT.

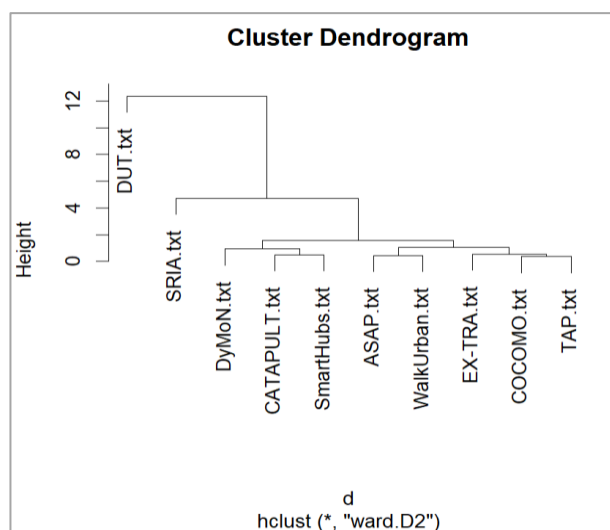


Figure 19: DUT & SRIA & 8 ENUAC project final reports texts (Height - ED between clusters).

The clustering result by 8 project final report, DUT and SRIA words is shown by dendrogram (see Figure 20).

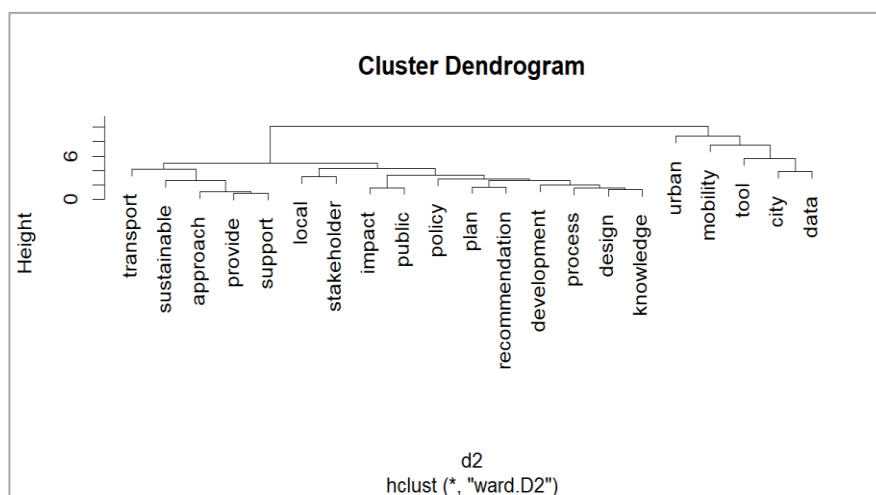


Figure 20. DUT & SRIA and 8 ENUAC projects final report words (Height - ED between clusters).

It can be inferred that words sharing the greatest similarity can be grouped together into the following clusters of similarity:

1. Cluster: Policy, recommendation & stakeholder;
2. Cluster: Sustainable, transport & support;
3. Cluster: Urban, city, mobility & tool.

The final reports emphasize stakeholder involvement, policy recommendations, and sustainability, indicating that the projects aim to leave practical, community-focused outcomes. The inclusion of terms like “data” and “tool” shows a practical focus on creating resources that can be used at the local level, supporting direct application beyond the projects themselves. The final reports remain distinct from the high-level strategic focus of DUT and SRIA, highlighting the applied nature of the ENUAC projects. Projects are implementing solutions and tools that align with strategic goals but are more focused on immediate, real-world impacts.

In summary, the final reports of the ENUAC projects reflect a shift toward concrete results, with a strong emphasis on community engagement, practical tools, and data-driven solutions. While the projects maintain alignment with the overarching goals of DUT and SRIA, their language and focus remain more grounded in operational and local applications.

1.4. Unique terms identification and text gap analysis

Unique terms identification and text gap analysis are valuable tools for understanding project documents, especially within a varied portfolio like ENUAC. These methods rely on TF-IDF (Term Frequency-Inverse Document Frequency) [6], a measure that highlights key terms in a document based on their frequency relative to other documents. **TF-IDF is particularly effective for identifying unique terms that stand out in specific documents, allowing a clearer view of each project’s distinct focus.**

Identifying unique terms highlights what makes each project distinct, helping clarify each project’s specific goals within the broader portfolio. For example, a project with “logistics” as a unique term likely has a different focus than those centred on “public transport” or “urban infrastructure.” By showing each project’s unique focus, this method makes it easier to understand what sets each project apart and can reveal potential collaboration opportunities. For instance, if one project focuses on “policy” and another on “stakeholder engagement”, they might complement each other’s work. Unique terms also help group projects by truly shared themes, not just generic terms.

Text gap analysis, on the other hand, finds topics that are missing or underrepresented across projects. By using a method like TF-IDF, it identifies important terms that appear rarely or not at all, highlighting possible gaps. For example, if “sustainability” is a key goal in urban planning but shows up infrequently, it suggests the need for more environmentally focused projects. Text gap analysis guides future project planning by identifying these gaps, helping ensure the portfolio aligns with strategic goals like sustainability or inclusivity.

Together, unique terms identification and text gap analysis provide a complete view of the project portfolio. Unique terms show each project’s focus, while text gap analysis identifies missing themes. This combined approach helps ensure that the portfolio clearly understands each project’s value, addresses any gaps, and aligns with key priorities in urban mobility.

Figure 21 displays the top 20 unique words for both DUT and SRIA. Based on the analysis, **the gap between DUT, SRIA and 15 ENUAC projects proposals** can be observed in the following words: 15minC (the 15-Minute City Transition Pathway), CUE (the Circular Urban Economy), PED (the Positive Energy Districts Transition Pathway), **energy**, portfolio, regenerative, partnership, **ecosystem**, **neutral**, **climate**, **transformation**, **economy**, dilemma, priority, **transition**, joint, strategic, instrument, alignment.

The gap is illustrated by comparing the unique words in DUT/SRIA with those unique to each project. By analyzing each project individually to obtain a list of unique words, a clear difference emerges among the top 20 words. This highlights the disparity between the projects and DUT/SRIA priorities.

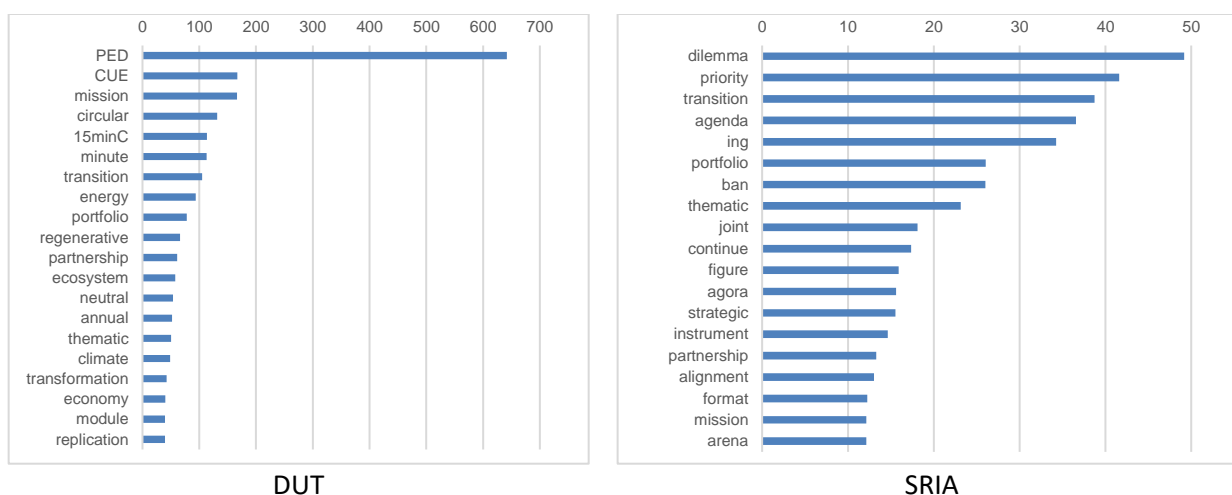


Figure 21: The top 20 unique words for DUT and SRIA, based on 15 ENUAC projects proposals (measured by TF-IDF).

Figure 22 displays the top 20 unique words for both DUT and SRIA. Based on the analysis, **the gap between DUT, SRIA and 15 ENUAC projects 1st & 2nd year report** can be observed in the following words: PED (the Positive Energy Districts Transition Pathway), **energy**, CUE (the Circular Urban Economy), **circular**, **climate**, 15minC (the 15-Minute City Transition Pathway), **neutral**, **economy**, **ecosystem**, partnership, regenerative, green, transformation, innovation, global. The gap is shown by comparing unique words in DUT/SRIA with those specific to each project.

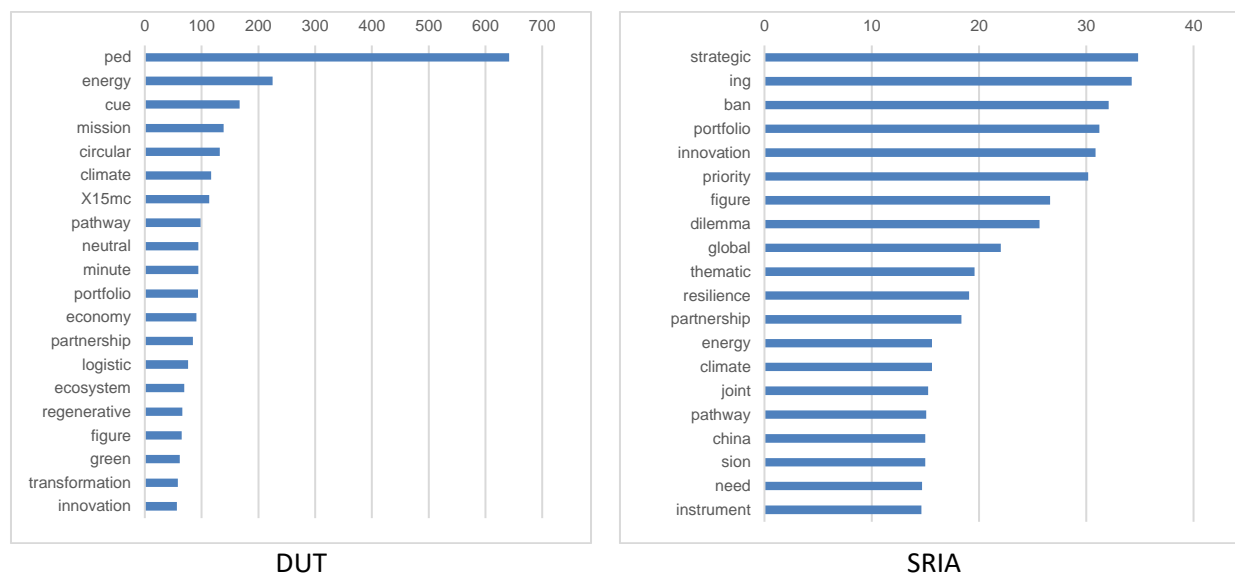


Figure 22: The top 20 unique words for DUT and SRIA, based on 15 ENUAC projects 1st & 2nd year report (measured by TF-IDF).

Figure 23 displays the top 10 unique words for both DUT and SRIA. Based on the analysis, **the gap between DUT, SRIA and 8 ENUAC projects final report** can be observed in the following words: **energy and climate**. The gap is represented by comparing unique words of DUT/SRIA and words unique for each project. Analysing each project separately to acquire list of unique words results in clear difference between top 20 words, highlighting the gap between projects and DUT/SRIA calls.

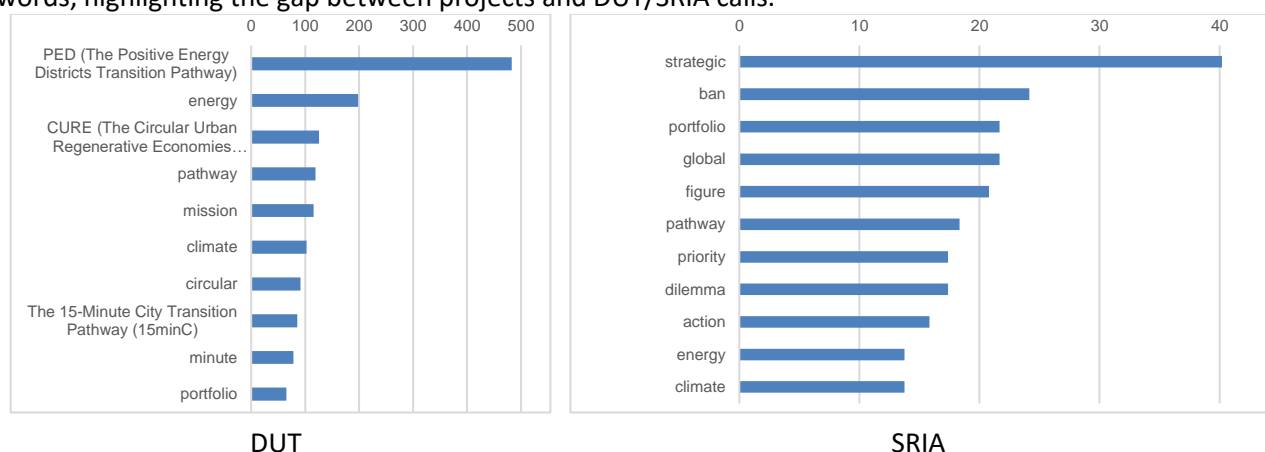
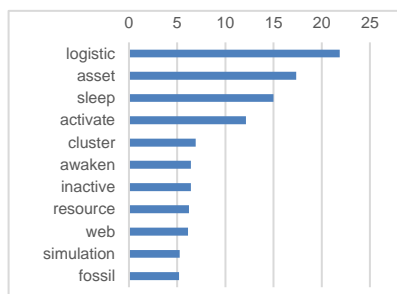


Figure 23: The top 10 unique words for DUT and SRIA, based on 8 ENUAC projects final report (measured by TF-IDF).

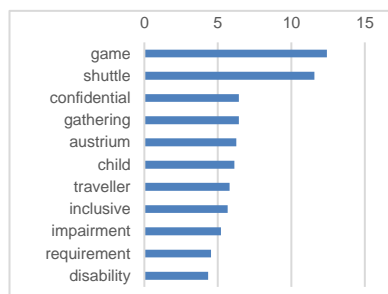
1.4.1 Uniqueness in project proposal's report text

The word clouds give the possibility to evaluate the most frequent words in each document and analysis of each document is performed separately. TF-IDF methods allow one to evaluate the most frequent words in corpus of documents; it means the analysis of 15 ENUAC project proposals was done together and the highest TF-IDF values show unique words in each project.

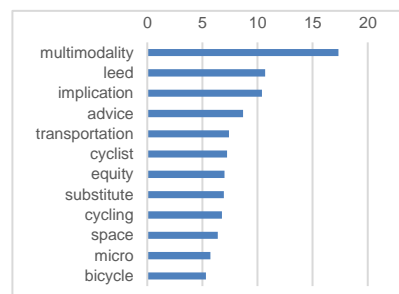
The 15 ENUAC projects were combined in one corpus and analysis was done for corpus of 15 documents. The top 10 unique words for 15 proposals of ENUAC projects are given in Figure 24.



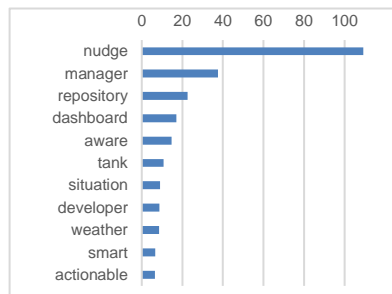
ASAP



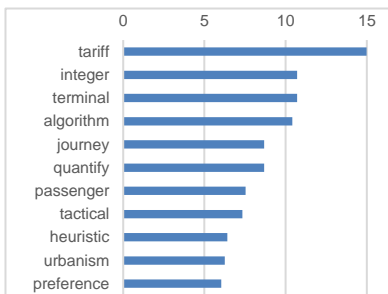
Catapult



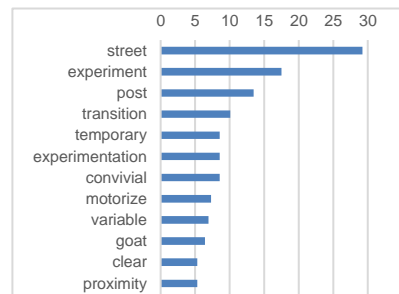
COCOMO: smm (shared micro-mobilities)



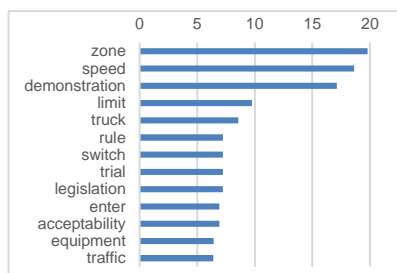
DyMoN



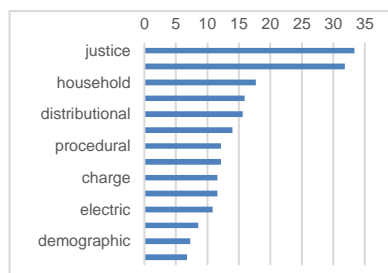
EASier



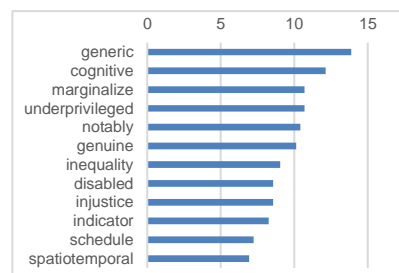
Ex-TRA



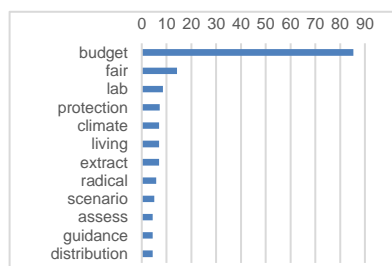
Geo-Sence



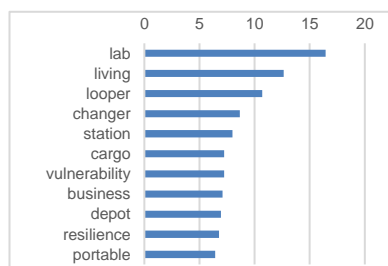
ITEM



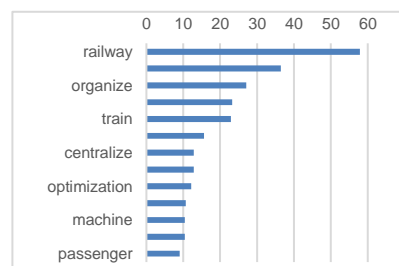
JUSTICE



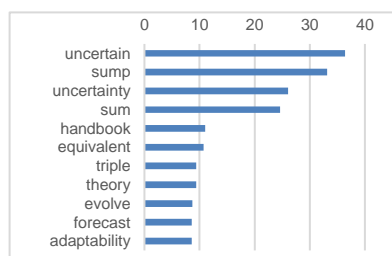
MyFairShare



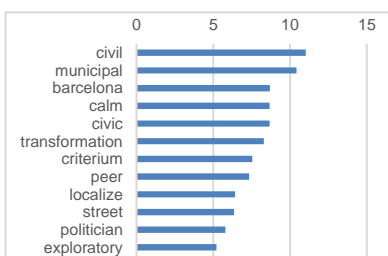
SmartHubs



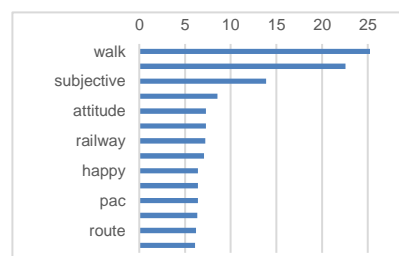
SortedMobility



TAP



TuneOurBlock



WalkUrban

Figure 24: The top 10 unique words for 15 ENUAC projects proposal texts (measured by TF-IDF).

1.4.2. Uniqueness in project 1st & 2nd year report text

The 15 ENUAC projects 1st & 2nd year report were combined in one corpus, and analysis was done for corpus of 15 documents. The top 10 unique words for 15 ENUAC projects 1st & 2nd year reports are given in Figure 25.

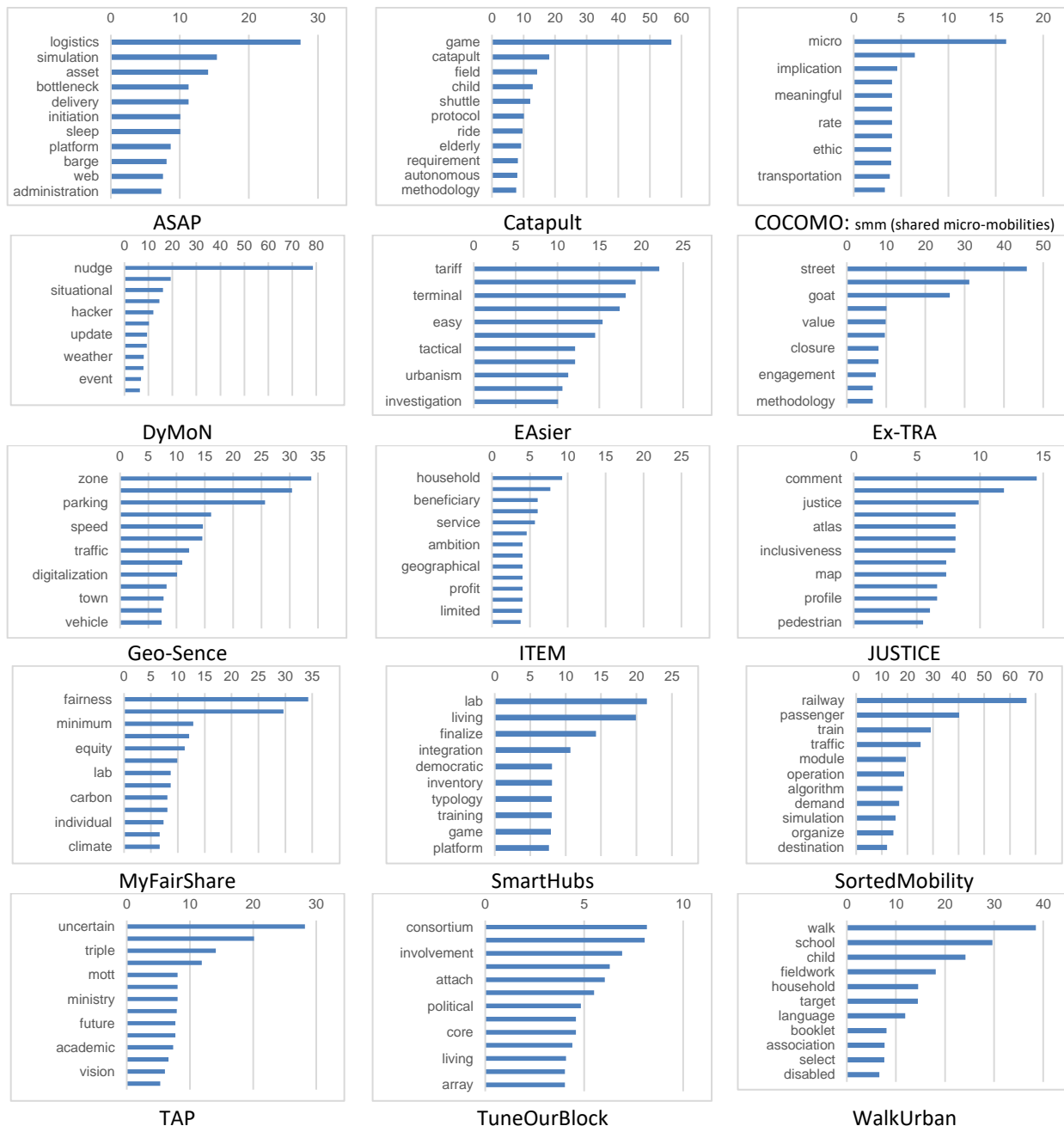


Figure 25: The top 10 unique words for 15 ENUAC projects 1st & 2nd year report texts (measured by TF-IDF).

1.4.3. Uniqueness in project final report text

The final report of 8 ENUAC projects were combined in one corpus, and analysis was done for the corpus of 8 documents. The top 10 unique words for the final text of 8 ENUAC projects are given in Figure 26.

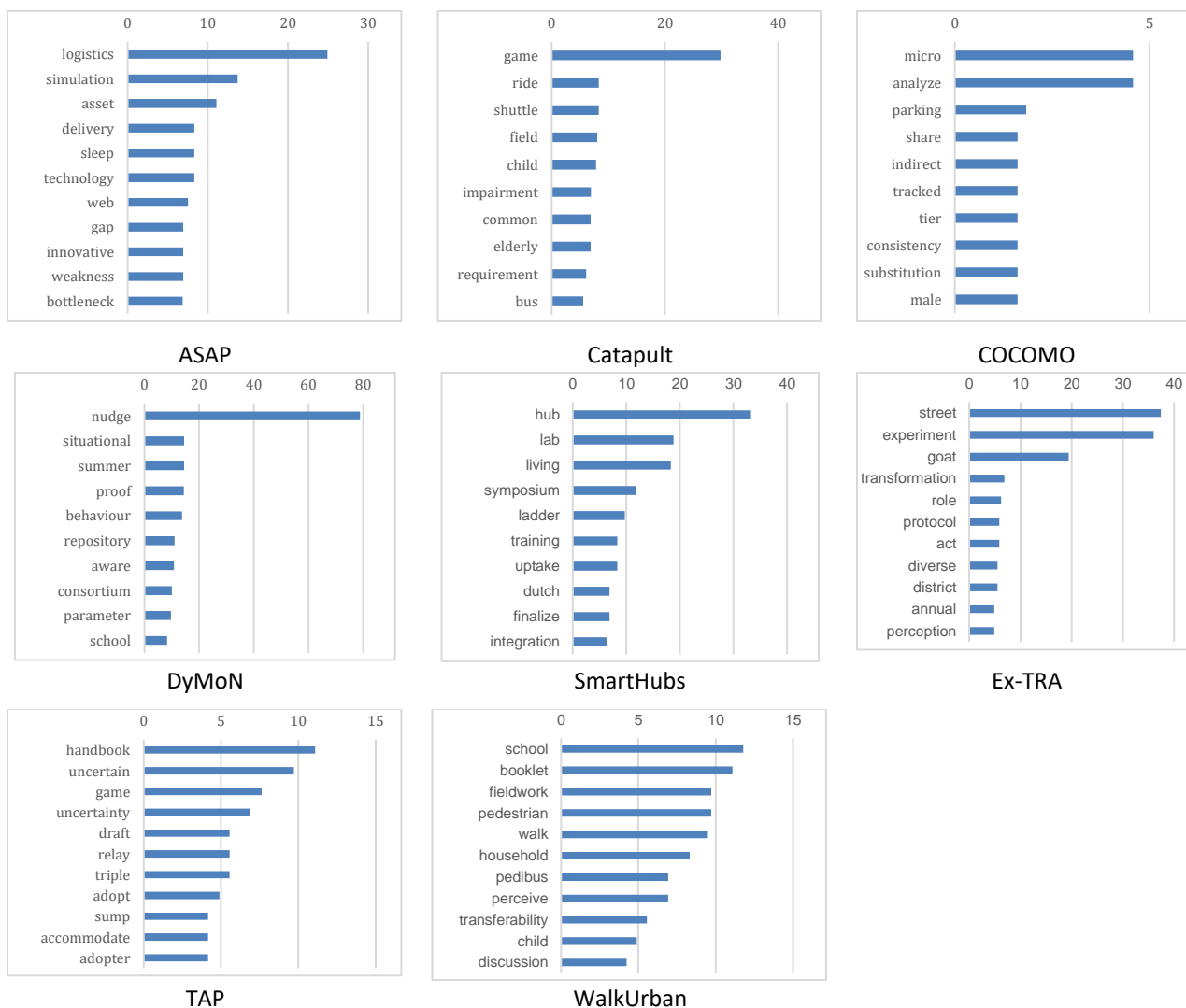


Figure 26: The top 10 unique words for 8 ENUAC projects final report texts (measured by TF-IDF).

2. Summary of textual analysis

Text-analytic methods can be used for document text analysis in different stages of documents evaluation starting from the project proposal and continuing with project mid-term and final reports. Text-analytic methods significantly reduce manual work compared to conventional methods and also enable us to interpret the initial information of the documents, understand the connections between them, and identify the unique topics. The analysis can also identify unique groups, showing which projects stand out with different perspectives or innovative ideas. This helps to spot projects that bring something new to the portfolio.

2.1. Conclusions

As the number of evaluated projects within the DUT programme rises, text analysis methods will facilitate faster and more scalable comparative assessments among them compared to traditional approaches (see Table 1).

Table 1: Summary of proposed text-analytic methods.

Text-analytic methods	Evaluated for	Applied for	Recommended for analysis
Word frequency analysis	each document separate analysis	<ul style="list-style-type: none"> main topics and areas of focus give shift of projects developing (mid-term report) 	<ul style="list-style-type: none"> proposal mid-term report final report other documents
Text similarity analysis by cluster analysis	analysis of all documents (corpus of documents) together in corpus	<ul style="list-style-type: none"> how closely related words in the corpus to identify patterns and themes how closely related documents by text 	in document corpus <ul style="list-style-type: none"> all proposals all mid-term reports Final reports project documents with DUT and SRIA
TF-IDF	analysis of all documents (corpus of documents) together in corpus	<ul style="list-style-type: none"> is particularly effective for identifying unique terms that stand out in specific documents allowing a clearer view of each project's distinct focus. 	in document corpus <ul style="list-style-type: none"> all proposals all mid-term reports final reports project documents with DUT and SRIA

Text analytics methods can be used as an additional tool for project evaluation because these methods have various limitations that can impact the quality and accuracy of the results.

Word Frequency Analysis identifies the main topics and areas of focus within each document by isolating frequently occurring terms. This method is **particularly useful for assessing the evolution of project themes** across reports (from proposal through mid-term to final) and **for capturing shifts in focus over time**. However, word frequency analysis **does not account for the context or relationships between terms** and may ignore nuances due to stopwords and word order. Thus, while word frequency analysis offers a valuable overview, it **requires complementary methods** to interpret the deeper thematic relationships.

Text Similarity Analysis by Cluster Analysis groups projects based on textual similarity, **revealing patterns and themes across documents** by examining how closely related certain words and phrases are within a corpus. This clustering **highlights thematic connections between projects**, but **interpretability of clusters remains a challenge**, especially when clusters are formed without additional contextual information. Clusters alone **do not always convey why specific projects align**, necessitating supplementary insights from other methods, like word frequency analysis and TF-IDF, for a clearer understanding.

The word clouds give the possibility to evaluate the most frequent words in each document, whereas the **TF-IDF method** allows one to evaluate the most frequent words in corpus of documents, and the highest TF-IDF values show unique words in each project. **The TF-IDF method** has several limitations, particularly in terms of capturing semantic meaning, handling synonyms, and word order.

TF-IDF is **highly effective for identifying unique terms that stand out in specific documents**, allowing a deeper understanding of each project's distinct focus within the corpus. While this method excels in distinguishing unique terms, it also **has limitations, particularly with regard to semantic meaning, synonyms, and word order**, which can lead to gaps in understanding the full thematic context. TF-IDF, therefore, **benefits from integration with clustering and word frequency analysis** to create a comprehensive thematic map.

The combined application of these methods overcomes individual limitations and enhances the overall quality of text analytics. For instance:

- Word clouds (based on word frequency analysis) provide an accessible visual foundation, highlighting prominent terms, which can serve as a preliminary step in understanding clusters.
- Cluster analysis can then use these prominent terms to establish structured groupings, clarifying how projects relate to one another.
- TF-IDF can further refine each cluster by identifying unique, cluster-specific terms, providing specificity on what sets each group of projects apart within the larger corpus.

This layered approach enables a more robust evaluation, supporting decisions regarding project alignment with DUT/SRIA goals and uncovering both commonalities and unique. While text analytics provides valuable insights for project evaluation, it has inherent limitations. These methods:

- Do not capture full context or semantic meaning without additional interpretative steps.
- Struggle with interpretability in cases where clustering lacks direct thematic support from word frequency or unique term analysis.
- Require complementary qualitative review (such as abstract analysis) to fully interpret and validate findings.

Thus, while these methods can significantly enhance project evaluation efficiency, they are best used in conjunction with human expertise and traditional evaluation techniques to ensure that the analysis aligns with strategic objectives.

2.2. Application

The flow of textual analysis during a research program can be specified with steps like topic setting, call development, proposal preparation phase, proposal evaluation phase, project start (with initial documentation), midterm and final evaluation, and future planning (see Figure 27). In details, each step can be described as follows:

1. **Topic Setting:** the goal is to identify emerging themes, priority areas, and gaps in the field that align with strategic objectives (e.g., of SRIA/DUT). The expected result is the definition of thematic clusters and gaps for setting the program's focus. These clusters and gaps define the thematic focus for Call Development, ensuring the call targets relevant and strategically aligned themes identified in this phase. The following methods can be applied:
 - *Word Frequency Analysis (WFA)* of recent publications, previous project summaries, or reports to pinpoint common and trending terms.
 - *TF-IDF* on thematic documents to uncover unique terms and underrepresented topics.
 - *Cluster Analysis (CLA)* or *Clustering* to group related topics and identify broad thematic areas.
2. **Call Development:** the goal is to develop a call for proposals that aligns with strategic objectives and highlights desired themes. The expected result is framework for the language and structure of the call document, ensuring alignment with strategic goals. The thematic language and structure set in this phase provide benchmarks for Proposal Phase screening, ensuring incoming proposals align with the program's focus. The methods *TF-IDF* and *Word Frequency Analysis* on strategic documents to extract key terms and themes to be emphasized in the call can be applied.
3. **Proposal Phase:** the goal is to screen incoming proposals for thematic alignment and identify potential project clusters. The expected result is an initial understanding of which proposals best align with call objectives and identifies potential clusters for further review.

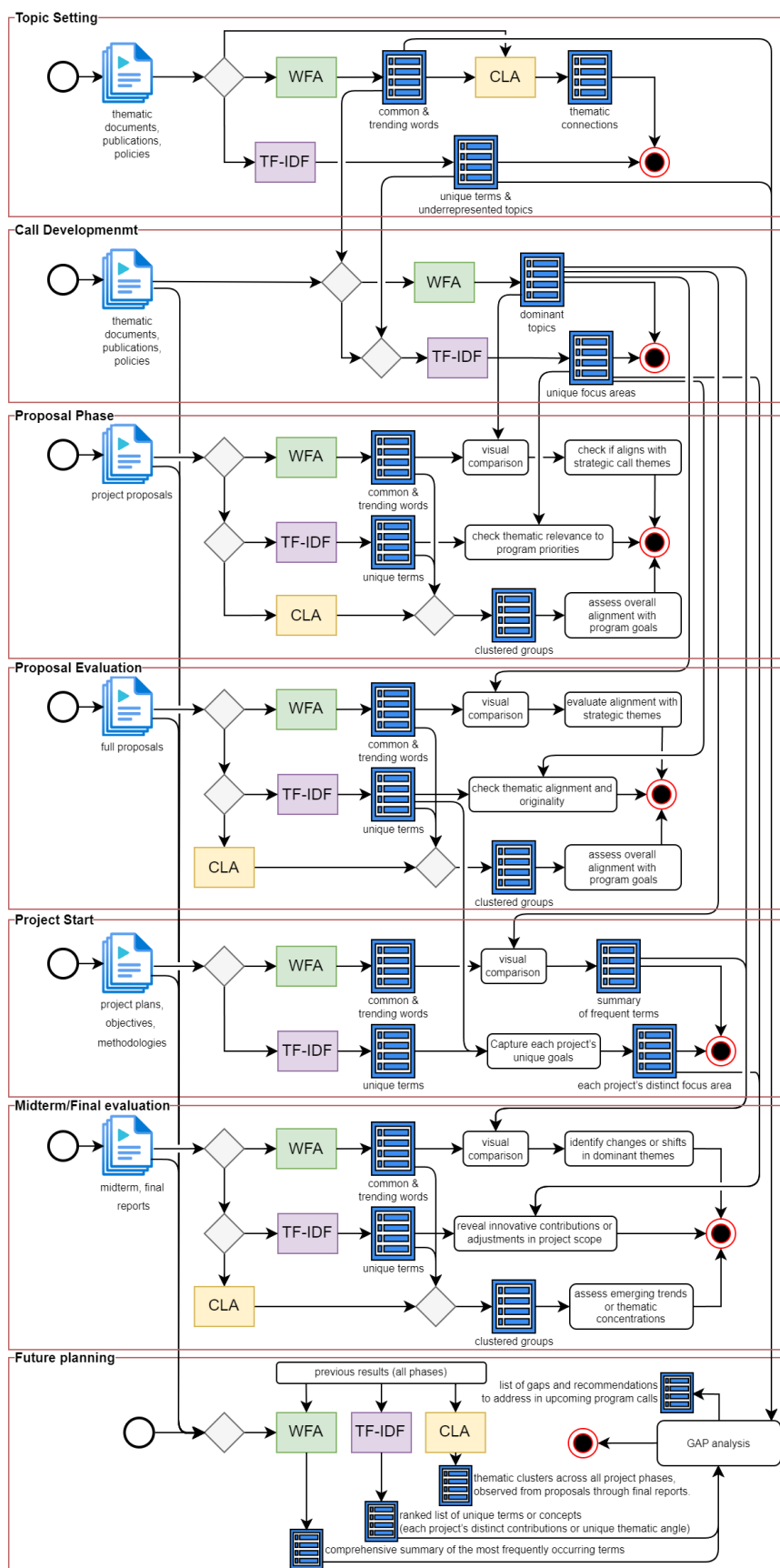


Figure 27: The flow of textual analysis application during a research program.

The identified clusters and thematic alignment from this phase inform Proposal Evaluation, helping evaluators prioritize proposals that best match strategic goals and unique contributions. The following methods can be applied:

- *Word Clouds* as a preliminary step to visualize prominent terms across proposals.
- *TF-IDF* to identify unique terms in each proposal, highlighting distinctive focus areas.
- *Cluster Analysis* to group proposals with similar themes and assess overall alignment with call priorities.

4. **Proposal Evaluation:** the goal is to evaluate and rank proposals based on compliance with strategic priorities and uniqueness. This should help evaluators prioritize proposals that align with strategic themes, meet quality standards, and provide unique contributions. Ranked proposals and identified themes are carried forward to Project Start, where initial thematic baselines are set based on each project's approved focus areas. The following methods can be applied:

- *TF-IDF* and *Word Frequency Analysis* to examine term frequency and compliance with call terms.
- *Cluster Analysis* to reinforce thematic grouping and help reviewers understand which proposals complement each other.

5. **Project Start:** the goal is to set baseline themes and objectives to track progress throughout the project lifecycle. The expected result is a thematic snapshot that will be used for comparison in later stages. Baseline themes and unique contributions provide a reference for Midterm/Final Evaluation to track thematic evolution and alignment with initial objectives. The following methods can be applied:

- *Word Frequency Analysis* on initial project documents to create a baseline for thematic focus.
- *TF-IDF* to capture unique goals or approaches specific to each project.

6. **Midterm Evaluation:** the goal is to assess the progression of each project and identify any shifts in focus. This should help evaluators to monitor project alignment with initial objectives and strategic goals, while noting new developments or gaps. The following methods can be applied:

- *Word Frequency Analysis* to track changes in dominant themes compared to the proposal phase.
- *TF-IDF* to highlight any new or emerging terms in midterm reports.
- *Cluster Analysis* to identify evolving thematic groups or deviations from original clusters.

7. **Final Report Evaluation:** the goal is to evaluate final outcomes and thematic contributions of each project, noting compliance and unique impacts. The expected result is a comprehensive view of the project portfolio's thematic alignment, identifying lasting contributions and areas for future exploration. Final outcomes provide cumulative insights into successful themes and potential gaps, which are essential for Future Planning and continuous improvement. The following methods can be applied:

- *Word Frequency Analysis* and *TF-IDF* on final reports to capture final themes and unique project outputs.
- *Cluster Analysis* to assess whether projects align within intended clusters or have introduced novel themes.

8. **Future Planning:** the goal is to synthesize learnings and refine strategic focus for future calls. The expected result is a feedback loop for continuous improvement, ensuring that future programs and calls are informed by insights from completed projects. The following methods can be applied:

- Integrated Analysis (*TF-IDF*, *Word Frequency*, *Cluster Analysis*) across all project phases to extract trends, successful themes, and gaps.

3. Comparison of D1.4 to D2.2 and D2.3

Both D1.4 and D2.2 & D2.3 look at the same 15 ENUAC projects within the DUT (Driving Urban Transitions) program, which aims to improve urban accessibility and connectivity. While they share similar goals that align with DUT's mission for sustainable urban development, their approaches are very different, making direct comparison difficult.

D1.4 takes a practical approach, focusing on how projects are carried out, the challenges they face, and immediate research needs. It gathers information from surveys with project stakeholders and provides insights into real-life applications like pilot programs and tool development. D1.4 highlights the importance of partnerships with local governments and industry, and it identifies new research areas such as public engagement and scaling up innovations, which align with DUT's goal of driving real urban change.

On the other hand, D2.2 & D2.3 take a broader, strategic approach. They assess how projects fit within DUT's larger goals, looking at areas like project vision, accessibility, and stakeholder involvement. These documents also offer recommendations to improve future projects, such as balancing leadership roles among different partners and clarifying project goals, supporting DUT's goal of creating lasting impact in urban transitions. Since D1.4 focuses on practical details and D2.2 & D2.3 address overall project structure, a direct comparison isn't possible. D1.4 gives immediate insights, while D2.2 & D2.3 look at bigger-picture issues and recommend adjustments for the future.

D1.4, D2.2, and D2.3 all agree on the importance of working closely with local stakeholders and creating lasting, adaptable solutions for urban projects. They emphasize that involving groups like city officials and communities is essential to ensure that projects are practical and relevant to real-world needs. The reports also stress the need to develop flexible tools and methods that can be applied to other cities, enhancing the projects' overall impact. In summary, D1.4 and D2.2 & D2.3 offer different but complementary views of the ENUAC projects. Together, they provide both practical advice and strategic guidance to help future DUT projects be effective and aligned with broader urban goals.

4. Conclusions and recommendations

The following conclusions and recommendations are formulated in relation to synthesized results for the development of future DUT programme:

1. In WP4 the main themes are identified for **DUT, SRIA & ENUAC projects** depending on the year of implementation:
 - for the **project proposals** there are three main clusters:
 - 1st cluster concept implementation **strategy implementation & network infrastructure**;
 - 2nd cluster includes **transport accessibility & policy**;
 - 3rd cluster includes **urban city mobility**.
 - for the **project 1st & 2nd year report** there are three main clusters:
 - 1st cluster includes **strategy implementation & network infrastructure & accessibility**;
 - 2nd cluster includes **transport policy**;
 - 3rd cluster includes **urban city mobility & traffic**.
 - for the **project final report** there are three main clusters:
 - 1st cluster includes **policy, recommendation & stakeholder**;
 - 2nd cluster includes **sustainable, transport & support**;
 - 3rd cluster includes **urban, city, mobility & tool**.
2. The **compliance or main themes** for **DUT & SRIA & ENUAC projects** are:

- **Strategy and model development, implementation & network infrastructure;**
 - **Transport accessibility & policy;**
 - **Urban city mobility.**
3. The most **common words** for **DUT and SRIA** are **urban, city, innovation, energy & transition** and for **ENUAC projects** are **mobility & city**. Comparison with DUT Roadmap text & word co-occurrence analysis shows that **energy & transition** are words with a lower co-occurrence of two adjacent terms in a text corpus.
 4. The similar result is found in the identification of **unique terms** for DUT&SRIA are **energy, ecosystem & climate**. Despite DUT and SRIA's focus on urban transition and sustainability, terms like “**energy**”, “**climate**”, and “**ecosystem**” are less common in ENUAC project documentation. This indicates a need for more environmentally-focused projects or stronger integration of sustainability themes in urban mobility and accessibility. Similarly, terms related to circular and regenerative economies – such as “**Circular Urban Economy**” (CUE) and “**regenerative**” – are also underrepresented. This points to a need for projects that prioritize resource efficiency, waste reduction, and circular practices within urban systems, aligning with the strategic goals of a circular urban economy.
 5. Text-analytic methods provide a powerful framework for both pre-funding and post-funding evaluation stages in project analysis. Prior to any funding decision, these methods enable a rapid, data-driven analysis of project proposals against call documents. This pre-funding stage allows evaluators to efficiently assess proposals for compliance with strategic goals, identify thematic trends (e.g., research topics, methodologies, geographic focus), and detect potential biases or gaps in the proposed research areas. In the post-funding phase, text-analytic methods continue to be valuable for evaluating midterm and final reports. By comparing these reports to initial project goals and the DUT/SRIA frameworks, evaluators can check if the project objectives are being met and if any new research gaps or emerging themes have developed over time.
 6. The following key points from textual analysis can be stated:
 - Methods should be applied in a complex, integrated manner at each phase, ensuring a balance between thematic identification, compliance assessment, and uniqueness evaluation.
 - Early stages (topic setting, call development) focus on setting strategic priorities using foundational analyses.
 - Proposal and evaluation phases emphasize compliance and thematic clustering to align projects with program goals.
 - Midterm and final reports assess progress and contributions, providing data for future strategic adjustments.
 7. The WP1 analysis centred on developing shared visions and assessing project impacts through pilot and testbed implementation, tool development and solution testing, digital and analog serious games, challenges for maximising tool impact. It also involved identifying lessons learned, evaluation outcomes, and future directions for research and practice. The primary objectives of the 15 ENUAC projects fall into four key themes:
 - **sustainability and efficiency** in urban mobility;
 - **innovative technology solutions** to urban mobility challenges;
 - **activation of spaces and infrastructure** towards alternative urban mobility and increased sustainability;
 - **inclusivity and accessibility** highlighting dimensions of equity and justice with regards to urban mobility.

8. The projects were put into practice through pilot programs and testing sites where solutions and tools were tried out in real city environments. They used real-life labs as spaces for evaluations and tests. The projects also created and tested tools like software applications, digital and physical educational games, tools to engage people, data collection tools, free guidelines, and prototypes. These educational games provided interactive ways for learning, training, or simulating real-world situations, involving stakeholders both online and in person.
9. The projects faced several challenges in getting the most out of their tools. Limited time and resources made it hard for university partners to fully test and apply the tools in wider settings. Sharing knowledge and solutions between different cities was difficult because insights were often specific to local areas. Also, keeping long-term benefits requires careful checking to make sure that early successes lead to lasting positive results without causing unwanted side effects.
10. New research questions and topics have been found in WP1 to guide future work in urban accessibility and connectivity. A main focus is combining digital tools with public involvement to make them easy to use and accessible to everyone. Another important area is expanding successful urban innovations to different sizes and various cities around the world, adapting them to different situations. Setting up support systems for living labs is also essential, including the legal, financial, and technical setups needed to keep them running after the initial tests:
 - **integration of digital tools and public participation;**
 - **scalability and transferability of urban innovations;**
 - **supportive infrastructures for Living Labs;**
 - **community and stakeholder engagement;**
 - **socio-technical transitions and policy impacts;**
 - **long-term sustainability and funding models;**
 - **policy and public perception challenges;**
 - **emerging questions from field implementations;**
 - **data and impact.**
11. The WP2 analysis found gaps in research and implementation in the ENUAC projects. One gap is a power imbalance, that research organizations dominate decision-making, even though working together with local people is crucial for lasting impact. There's also an innovation bias - a tendency to support small new ideas rather than challenging existing systems. Additionally, the projects often don't explain key concepts or deal with the challenges of applying solutions in different contexts, showing a lack of focus on understanding and sharing knowledge.
12. WP2 recommends for DUT to explore new approaches to ensuring that local stakeholders are actively involved in project formulation, execution, decision-making, and afterlife, and issue funding calls that accommodate a broader spectrum of project approaches in terms of experimentation and participation. The calls should also require applicants to better clarify their visions, theories of change, conceptual frameworks, and impact strategies.

5. References

1. JPI Urban Europe's EN-UAC **Knowledge Hub ACUTE** (*Accessibility and Connectivity knowledge hub for Urban Transformation in Europe*) – a platform that enables exchange of knowledge, expertise and results in urban accessibility and connectivity from various actors. <https://jpi-urbaneurope.eu/news/the-knowledge-hub-a-platform-for-gathering-knowledge-and-experiences-for-a-sustainable-city/>
2. JPI Urban Europe's programme **Driving Urban Transitions to a Sustainable Future (DUT)**. <https://dutpartnership.eu>
3. **D1.4 Final showcase of regional, national, European and international projects, ideas, initiatives on the topic of accessibility and connectivity – the ENUAC projects**. JPI Urban Europe's EN-UAC Knowledge Hub ACUTE (*Accessibility and Connectivity knowledge hub for Urban Transformation in Europe*) deliverable report, 2024.
4. **D2.2 Analytical framework and methodology**. JPI Urban Europe's EN-UAC Knowledge Hub ACUTE (*Accessibility and Connectivity knowledge hub for Urban Transformation in Europe*) deliverable report, 2024.
5. **D2.3 Research synthesis**. JPI Urban Europe's EN-UAC Knowledge Hub ACUTE (*Accessibility and Connectivity knowledge hub for Urban Transformation in Europe*) deliverable report, 2024.
6. Robinson, D., Silge, J. (2017). *Text Mining with R: A Tidy Approach*. United States: O'Reilly Media.
7. JPI Urban Europe's **ERA-NET Urban Accessibility and Connectivity (EN-UAC)** partnership. <https://www.era-learn.eu/network-information/networks/en-uac>
8. JPI Urban Europe on **Basecamp** - a space dedicated for exchanges with to with the first 15 ENUAC Co-funded projects coordinators. <https://3.basecamp.com/3660243/projects>
9. JPI Urban Europe's **Strategic Research and Innovation Agenda**, SRIA 2.0. <https://jpi-urbaneurope.eu/about/sria/sria-2-0/>
10. The Open Graph Viz Platform, <https://gephi.org/>
11. **Arhipova, I., Paura, L., Bumanis, N., Vitols, G., Salajevs, V., Erglis, A., Berzins, G., Ansonskā, E.** (2024) *Applying text analytics methodology to analyze project reports*. Proceedings of the 6th International Conference on Finance, Economics, Management and IT Business (FEMIB 2024), pp. 89–95. [Scopus - Document details - Applying Text Analytics Methodology to Analyze Project Reports](#)
12. Berry, Michael W., and Jacob Kogan. *Text Mining: Classification, Clustering, and Applications*. Hoboken, NJ: Wiley-Interscience, 2010.