



# A semiautomated framework Leveraging NLP for Skill Identification and Talent Management of the Acquisition Workforce in the Department of Defense

Dr. Jose Emmanuel Ramirez-Marquez  
Stevens Institute of Technology

Mr. Garry Shafovaloff  
Defense Acquisition University

Dr. Dinesh Verma  
Stevens Institute of Technology

Mark Krzysko  
Office of the Assistant Secretary of Defense

# Objective

By automating skill identification via NLP, the DoD can promptly identify areas of expertise and allocate resources accordingly, ensuring optimal deployment of personnel with the requisite skills to areas of utmost need. This streamlined process enhances resource allocation efficiency and augments readiness by mitigating skill gaps and elevating overall force preparedness.



**Skill Identification:** Utilize NLP algorithms to analyze reports provided by the sponsor, to automatically identify critical skills within the DoD workforce.



**Skill Mapping:** Develop maps of identified skills to specific job roles and career pathways within the DoD, facilitating strategic talent management and career development initiatives.

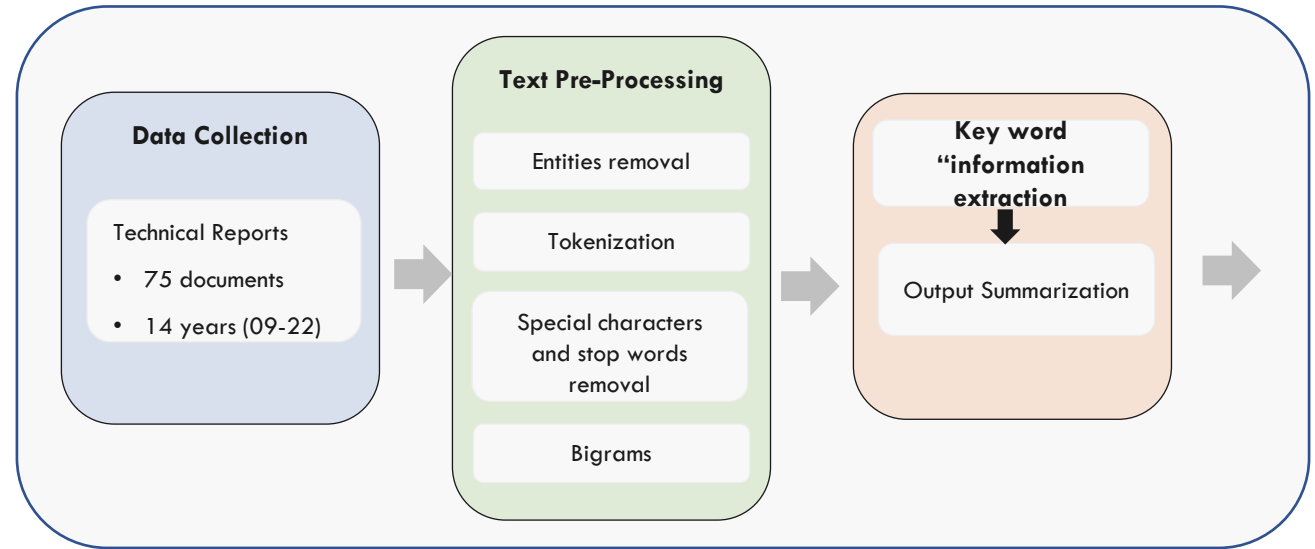


**How?**

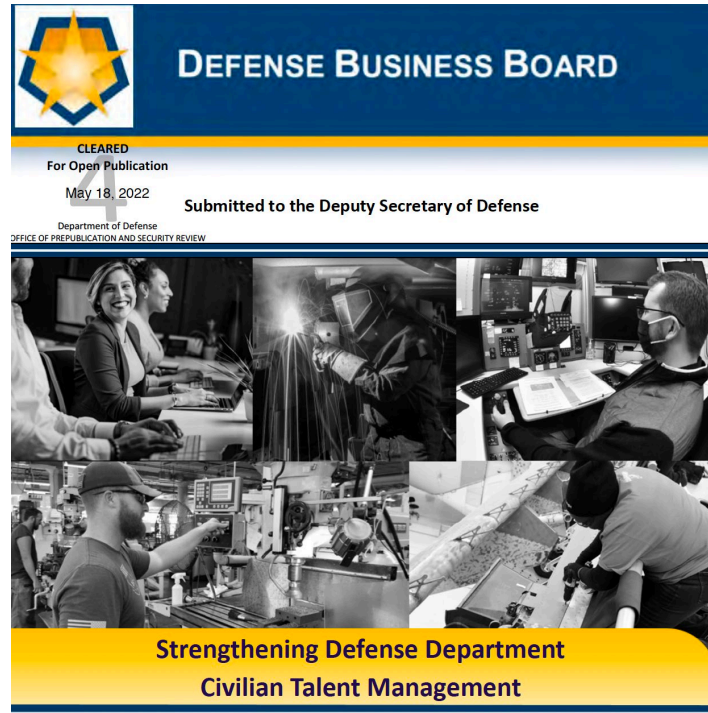


Analyzing technical reports related to civilian and DOD workforce skills

# Framework Phase 1



# Example 1 of framework implementation Phase 1



Input: Report (95 structured pages)

DBB 2022 - Strengthening DoD Civilian Talent Management  
The Department of Defense (DoD or "the Department") has difficulty attracting, recruiting, and hiring for critical skill sets in emerging technologies. Compounding this challenge, the DoD struggles to identify and track the skill sets that it already has in its civilian workforce and to match them to the appropriate jobs. Specifically, the Subcommittee will:

- Provide case studies and distill best practices from relevant private sector companies on how private industry projects forward the number and types of skills they will need in the future;
- Review the Department's current civilian workforce planning methods, identify gaps in best practices, and recommend changes leveraging on private sector practice;
- Review the Department's existing approaches to identifying and categorizing worker skill sets and tracking them over time.

Provide comparison examples of best practices from private industry or other public entities, and identify the laws, policies, or practices that inhibit implementation within the Department;

- Review the Department's approach to matching worker skill sets to the needs of particular jobs or career fields and identify practices that impede effective matching of employee skills to jobs.

The Board has physical storage capability and electronic storage and communications capability on both the unclassified and classified networks to support receipt of material up to the Secret level.

- This report shall include:
  - o How private industry projects the number and types of skills they will need in the future;
  - o The Department's current civilian workforce planning methods;
  - o The Department's existing approaches to identifying and categorizing worker skill sets and tracking them over time, and identifying the laws, policies, or practices that inhibit implementation within the Department;
  - o The Department's approach to matching worker skill sets to the needs of particular jobs or career fields and identify practices that impede effective matching of employee skills to jobs;
  - o Changes to statutory requirements that inhibit the Department's ability to reskill its civilian workforce;
  - o Case studies of large companies that structured successful reskilling/upskilling programs either enterprise-wide or within a major sector.

44 Subcommittee  
DBB Members

Output: Keyword information  
Extraction (7 unstructured pages)

```
[81]: from PyPDF2 import PdfReader
      from summarizer import Summarizer

      def extract_text_from_pdf(path):
          with open(path, 'rb') as file:
              pdf_reader = PdfReader(file)
              text = ''
              for page in pdf_reader.pages:
                  text += page.extract_text()
              return text

      def summarize_text(text, summary_length):
          summarizer = Summarizer()
          summarized_text = summarizer(text, num_sentences=summary_length)
          return summarized_text

      # Path to your PDF file
      pdf_path = '/Users/Pollac/Desktop/Tests/Test2.pdf'

      # Extract text from the PDF
      document_text = extract_text_from_pdf(pdf_path)

      # Set the desired summary length (in sentences)
      summary_length = 100

      # Summarize the document
      summary = summarize_text(document_text, summary_length)

      # Print the summary
      print(summary)

      Some weights of the model checkpoint at bert-large-uncased were not used
```

Summarize Output (1 page)

# Framework Output

1. Talent Acquisition and Recruitment: Skills in attracting and recruiting professionals with critical skill sets in emerging technologies. This includes expertise in sourcing candidates, conducting interviews, assessing qualifications, and employing effective recruitment strategies.

2. Workforce Planning: Skills in strategic workforce planning to anticipate and align human capital needs with organizational goals. This involves analyzing current and future skill requirements, identifying gaps, and developing plans to address those gaps through recruitment, training, or other talent management initiatives.

3. Skill Set Identification and Tracking: Competence in identifying and categorizing worker skill sets, as well as establishing systems to track and update these skills over time. This includes using technology and data analysis to monitor skill inventories, assess skill gaps, and ensure accurate matching of employee skills to job requirements.

4. Comparative Analysis and Benchmarking: Skills in conducting comparative analysis between the DoD's talent management practices and those of private sector companies or other public entities. This entails identifying best practices, gaps, and areas for improvement, and making recommendations for adapting private industry practices to enhance talent management in the DoD.

5. Skill Matching and Job Alignment: Abilities in matching worker skill sets to the needs of specific jobs or career fields within the DoD. This requires understanding the knowledge, experience, and competencies required for different roles and effectively aligning employees' skills to maximize their contributions and job satisfaction.

6. Change Management: Proficiency in managing change within the organization to facilitate the adoption of new talent management practices. This includes communication, stakeholder engagement, and developing strategies to address resistance or challenges related to implementing new approaches to talent management.

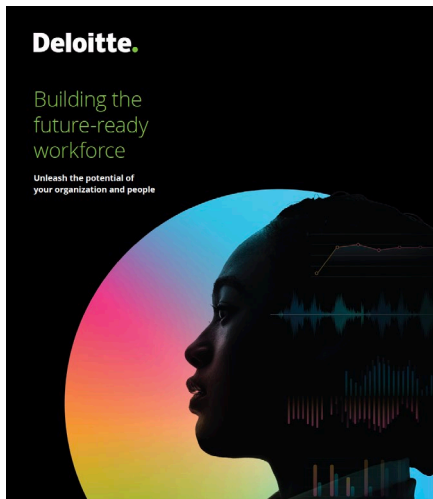
7. Data Management and Analysis: Competence in data management and analysis to support talent management decisions. This includes using technology platforms and data lakes to track and analyze job-related data, employee skills, and workforce trends, enabling informed decision-making and proactive planning.

8. Collaboration and Relationship Building: Skills in building partnerships and collaborative relationships with stakeholders both within and outside the organization. This includes fostering cooperation between different departments, leveraging external expertise, and engaging with private industry partners to exchange knowledge and best practices.

9. Talent Development and Upskilling: Abilities in designing and implementing talent development programs that enhance employees' skills and promote lifelong learning. This involves creating learning opportunities, providing access to training resources, and encouraging employees to expand their knowledge and expertise in line with organizational needs.

10. Knowledge of Emerging Technologies: Understanding and awareness of emerging technologies relevant to the defense sector. This includes staying abreast of advancements in areas such as cybersecurity, artificial intelligence, data analytics, robotics, and other emerging fields that impact DoD operations and require specialized skill sets.

# Example 2 of framework implementation



Input: Report (28 structured pages)

Deloitte 2021 - Building a future-ready military workforce  
Index  
In focus: Workforce, Skills, and Culture  
A Future of Warfighting publication by Deloitte Canada  
Source: Deloitte analysis.

Development and acquisition Resilient operations Workforce, skills, and culture Decision-making Repeatable, transparent acquisition processes

- National forces can move to a conflict, sustain and protect themselves, and apply force to an adversary
- Defined and accountable organizational culture in defense organizations
- Recruitment sufficient to maintain desired end-strength and contemporary skills
- Secure, reliable information systems
- Trustworthy data
- Timely data collection and analysis
- An understanding of policy and legal boundaries/permissions
- Ability to own and share technical data for select acquisition programs (e.g., via digital or model-based systems engineering)
- DevSecOps, Agile, or other iterative models of production used for select software development
- Mechanism for joint requirements development/coordination (e.g., JROC in the United States)
- Standards for joint interoperability of key systems
- Services have access to technical baseline data
- Flexible acquisition processes operating at the speed of technology
- Common operational standards for common tasks such as air support
- Ability to leverage other services' central military capabilities

Output: Keyword information Extraction (7 unstructured pages)

```
[81]: from PyPDF2 import PdfReader
      from summarizer import Summarizer

      def extract_text_from_pdf(path):
          with open(path, 'rb') as file:
              pdf_reader = PdfReader(file)
              text = ""
              for page in pdf_reader.pages:
                  text += page.extract_text()
              return text

      def summarize_text(text, summary_length):
          summarizer = Summarizer()
          summarized_text = summarizer(text, num_sentences=summary_length)
          return summarized_text

      # Path to your PDF file
      pdf_path = '/Users/Pollac/Desktop/Tests/Test2.pdf'

      # Extract text from the PDF
      document_text = extract_text_from_pdf(pdf_path)

      # Set the desired summary length (in sentences)
      summary_length = 100

      # Summarize the document
      summary = summarize_text(document_text, summary_length)

      # Print the summary
      print(summary)

Some weights of the model checkpoint at bert-large-uncased were not used
```

Summarize Output (1 page)

# Framework Output

The Deloitte 2021 report titled "Building a Future-Ready Military Workforce" emphasizes the need for militaries to take proactive measures in redesigning their personnel strategies, continuously assessing and adapting skills and culture to remain operationally ready now and in the future. The report highlights the importance of identifying and recruiting individuals with enduring capabilities that enable them to adapt and thrive regardless of mission changes.

Skills are defined as tactical knowledge specific to a particular context, while enduring capabilities are observable human attributes that endure and help individuals adapt and acquire new skills. The report suggests that military learning strategies should focus less on perfecting specific skills and more on training individuals to react and adapt to new situations. Interoperability with other nations and industries adds another dimension, requiring an understanding of allied skills and capabilities.

Recruitment is identified as a critical lever for building the future military workforce. Partnerships with industry, academia, and allies can provide access to external skills and expertise without the need for internal development. The report suggests considering strategies such as lateral entry and knowledge transfer from external occupations to broaden the talent pool and infuse the organization with industry best practices and new ways of thinking.

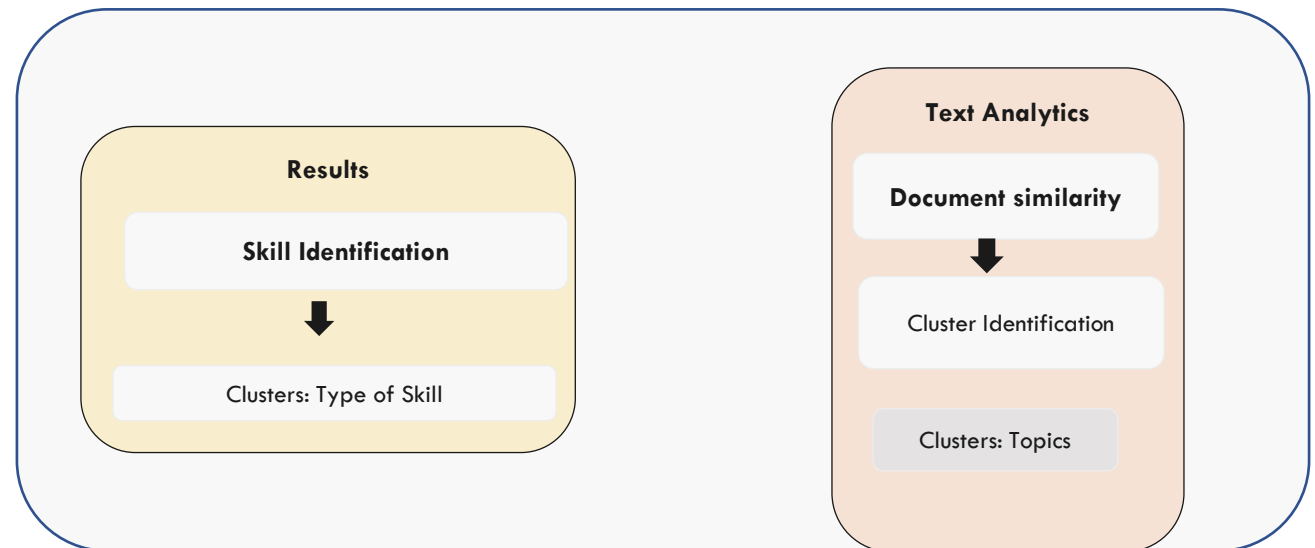
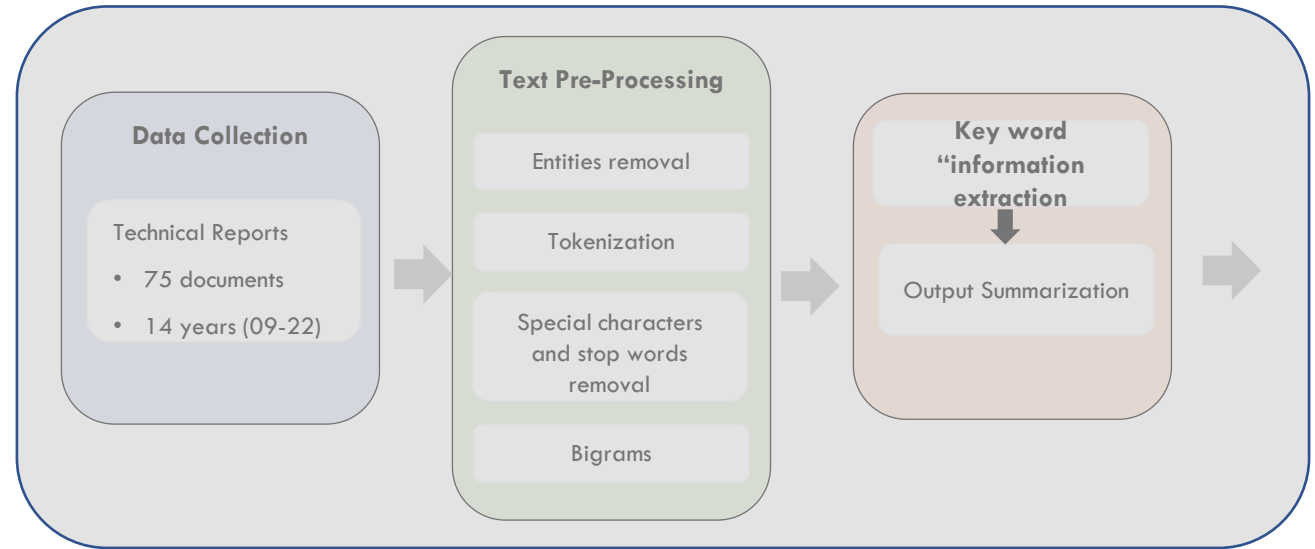
The evolving nature of skills and capabilities in defense necessitates the modernization of military learning strategies. This includes ensuring relevance, enabling servicemembers to acquire necessary skills when needed, and developing capabilities for growth and adaptability over time. The report highlights various tools available to support these strategies, such as real-time skills tracking and internal talent marketplaces.

To maintain operational readiness, the report emphasizes the importance of taking action today to redesign traditional military personnel strategies and build a future-ready workforce, skills, and culture. By embracing these measures, militaries can better prepare for the evolving challenges and ensure their readiness in an ever-changing operational environment.

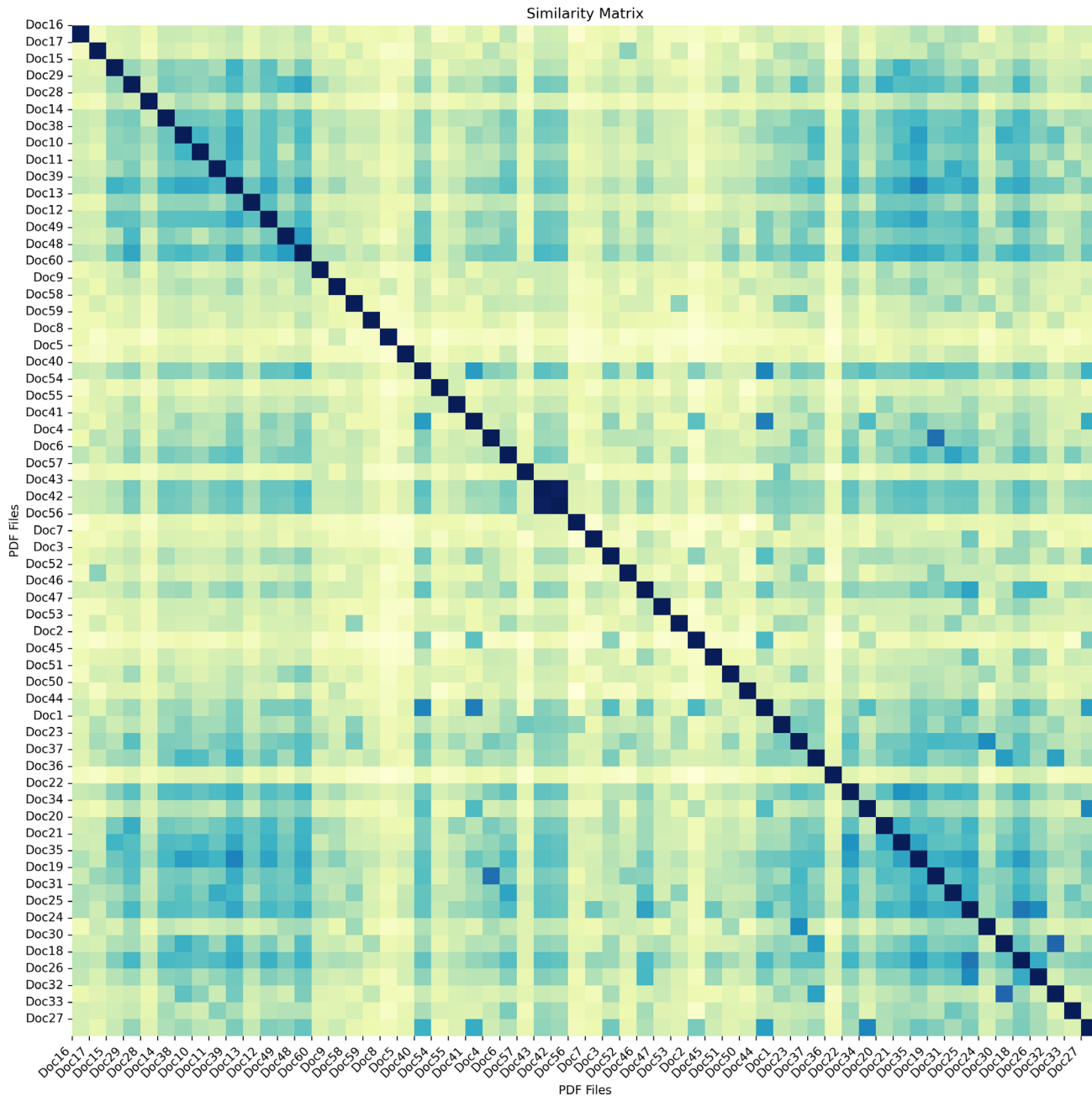
The following skills are relevant:

1. Critical thinking and analysis: The ability to evaluate information, analyze situations, and make informed decisions.
2. Problem-solving: The capacity to identify and solve complex problems in diverse contexts.
3. Active learning: The willingness and ability to acquire new knowledge and skills through ongoing learning and development.
4. Resilience: The ability to bounce back from setbacks, cope with challenges, and adapt to changing circumstances.
5. Flexibility: The capacity to adapt to new situations, tasks, and roles, demonstrating openness to change.

# Framework Phase 2:







Similarity Matrix:  
 Allows to identify clusters of documents that share similar contexts.  
 Helpful to determine the number of clusters

```

Clustering.ipynb  SimilarityMatrix.ipynb
Code

pdf_files = [file for file in os.listdir(documents_dir) if file.endswith('.pdf')]

# Read and preprocess the content of each PDF document
texts = []
for pdf_file in pdf_files:
    file_path = os.path.join(documents_dir, pdf_file)
    text = read_pdf(file_path)
    texts.append(text)

# Compute TF-IDF vectors for the documents
vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform(texts)

# Calculate the cosine similarity matrix
similarity_matrix = cosine_similarity(tfidf_matrix)

# Convert similarity matrix to DataFrame for better representation
similarity_df = pd.DataFrame(similarity_matrix, columns=pdf_files, index=pdf_files)

# Extract file names without extensions
pdf_names = [os.path.splitext(pdf_file)[0] for pdf_file in pdf_files]

# Set up the figure and axis
plt.figure(figsize=(16, 13))
ax = sns.heatmap(similarity_df, annot=False, cmap='YlGnBu', fmt=".2f")

# Customize the plot
ax.set_title("Similarity Matrix")
ax.set_xlabel("PDF Files")
ax.set_ylabel("PDF Files")
plt.xticks(range(len(pdf_names)), pdf_names, rotation=45, ha='right')
plt.yticks(range(len(pdf_names)), pdf_names, rotation=0)

# Save the plot as an image (JPEG or PNG)
output_image_path = os.path.join(documents_dir, "similarity_matrix.png")
plt.tight_layout()
plt.savefig(output_image_path, dpi=300)
plt.show()

# Output similarity matrix to a CSV file
similarity_df.to_csv(os.path.join(documents_dir, "similarity_matrix.csv"))

```

# Identify Document Clusters and provide output in terms of bi-grams

- Cluster 9 - Shared Bigrams: aspect life, skill item, skill framework, skill pattern, execut summar, scale internet, conclus studi, futur research, item clariti, method employ, comprehens set, test internet, develop reliabl, studi seek, interview conduct, et al, set reliabl, internet skill, differ popul, valuabl futur, type engag, find three, occup group, compar skill, base find, categori contain, studi aim, better understand, level measur, framework base, valid reliabl, internet use, systemat literatur, reliabl valid, polici evalu, report emphas, goe beyond, skill goe, aim develop, pattern relat, van deursen, measur internet, emphas import, five main, coher studi, import internet, reliabl measur, relat differ, inform navig, propos scale, literatur review, educ level, studi found, skill relat, categori oper, report focus, aspect internet, life outcom, impact variou, research better, use valid, holist view, across differ, measur skill, summar provid, examin extern, seek identifi, measur allow, reliabl cognit, identifi comprehens, digit literaci, age educ, inclus everyday, creativ mobil, skill scale, skill execut, skill level, main skill, differ aspect, three basic, methodolog studi, evalu propos, clariti coher, skill understand, develop internet, difficulti use, skill digit, consist skill, applic survey, literaci digit, propos skill, skill categori, technic aspect, user engag, investig skill, summar van, impact user, inclus effort, digit inclus, valid measur, life background, navig social, background focu, popul group, framework includ, level occup, view skill, develop set, outcom propos, valid report, found consist, relat gender, group conclus, valid internet, outcom studi, al measur, found item, employ investig, skill measur, understand relat, identifi studi, project report, basic method, focus measur, engag outcom, oper inform, inclus impact, specif skill, research intervent, propos framework, ensur accuraci, everyday life, contain specif, extern valid, conduct evalu, conduct systemat, valid compar, survey perform, accuraci understand, perform test, studi develop, variou aspect, evalu relat, item need, focu report, differ type, need revis, intervent polici, skill methodolog, use applic, social creativ, inclus propos, allow research, report examin, relat digit, engag impact, includ five, skill difficulti, deursen et, mobil skill, revis ensur, skill valuabl, understand digit, aspect studi, cognit interview, gender age, beyond technic, propos holist, understand extern, scale across, skill impact, review identifi, studi conduct, group found, framework propos, provid overview, overview project, impact digit Related Documents: - Doc59.pdf

```
Clustering.ipynb  SimilarityMatrix.ipynb  +
+  ✂  📄  ▶  ⏪  Code  📈

[4]: import os
import ssl
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
import numpy as np
from PyPDF2 import PdfReader

# Configure SSL context to bypass certificate verification for NLTK downloads
try:
    _create_unverified_https_context = ssl._create_unverified_context
except AttributeError:
    pass
else:
    ssl._create_default_https_context = _create_unverified_https_context

# Download necessary NLTK resources
nltk.download('punkt')
nltk.download('stopwords')

# Function to read text from PDF
def read_pdf(file_path):
    with open(file_path, 'rb') as file:
        pdf_reader = PdfReader(file)
        text = ""
        for page in pdf_reader.pages:
            text += page.extract_text()
        return text

# Function to preprocess the text and generate bigrams
def preprocess_text(text):
    # Tokenize the text into words
    words = nltk.word_tokenize(text.lower())
```

## Extract Meaning out of the cluster: AI-Generated

• **Cluster 9** - Shared Bigrams: aspect life, skill item, skill framework, skill pattern, execut sumari, scale internet, conclus studi, futur research, item clariti, method employ, comprehens set, test internet, develop reliabl, studi seek, interview conduct, et al, set reliabl, internet skill, differ popul, valuabl futur, type engag, find three, occup group, compar skill, base find, categori contain, studi aim, better understand, level measur, framework base, valid reliabl, internet use, systemat literatur, reliabl valid, polici evalu, report emphas, goe beyond, skill goe, aim develop, pattern relat, van deursen, measur internet, emphas import, five main, coher studi, import internet, reliabl measur, relat differ, inform navig, propos scale, literatur review, educ level, studi found, skill relat, categori oper, report focus, aspect internet, life outcom, impact variou, research better, use valid, holist view, across differ, measur skill, sumari provid, examin extern, seek identifi, measur allow, reliabl cognit, identifi comprehens, digit literaci, age educ, inclus everyday, creativ mobil, skill scale, skill execut, skill level, main skill, differ aspect, three basic, methodolog studi, evalu propos, clariti coher, skill understand, develop internet, difficulti use, skill digit, consist skill, applic survey, literaci digit, propos skill, skill categori, technic aspect, user engag, investig skill, sumari van, impact user, inclus effort, digit inclus, valid measur, life background, navig social, background focu, popul group, framework includ, level occup, view skill, develop set, outcom propos, valid report, found consist, relat gender, group conclus, valid internet, outcom studi, al measur, found item, employ investig, skill measur, understand relat, identifi studi, project report, basic method, focus measur, engag outcom, oper inform, inclus impact, specif skill, research intervent, propos framework, ensur accuraci, everyday life, contain specif, extern valid, conduct evalu, conduct systemat, valid compar, survey perform, accuraci understand, perform test, studi develop, variou aspect, evalu relat, item need, focu report, differ type, need revis, intervent polici, skill methodolog, use applic, social creativ, inclus propos, allow research, report examin, relat digit, engag impact, includ five, skill difficulti, deursen et, mobil skill, revis ensur, skill valuabl, understand digit, aspect studi, cognit interview, gender age, beyond technic, propos holist, understand extern, scale across, skill impact, review identifi, studi conduct, group found, framework propos, provid overview, overview project, impact digit

Related Documents: - **Doc59.pdf: Van Deursen et al. 2014 - Measuring Internet Skills**

### Skill Development:

#### 1. Internet Skills:

- Develop and measure internet skills across different populations.
- Enhance the ability to navigate and utilize the internet effectively.
- Improve digital literacy and inclusivity efforts for everyday life.

#### 2. Digital Skills:

- Develop digital skills, including technical aspects and creative mobile use.
- Increase proficiency in using digital tools and applications.
- Improve digital inclusivity and ensure accuracy in digital interactions.

#### 3. Cognitive Skills:

- Strengthen cognitive skills related to internet use and information navigation.
- Improve cognitive abilities for processing and comprehending digital information.
- Enhance skills in critically evaluating information from online sources.

#### 4. Analytical Skills:

- Develop analytical skills to examine and evaluate the impact of internet use on various aspects of life.
- Enhance the ability to conduct systematic literature reviews and perform valid comparative studies.
- Improve skills in conducting interviews and surveys to gather relevant data.

#### 5. Problem-Solving Skills:

- Enhance problem-solving skills related to digital challenges and difficulties in internet use.
- Develop strategies to address the needs of different user groups and improve inclusivity.

### Talent Development:

#### 1. Research Expertise:

- Attract and nurture talent with expertise in conducting research related to internet skills and digital literacy.
- Develop professionals who can create comprehensive and coherent studies on the impact of internet use.

### Capability Development:

#### 1. Framework Development:

- Develop a reliable and valid framework for measuring internet skills and digital literacy.
- Improve the framework to encompass a holistic view of internet-related skills.

#### 2. Inclusivity and Impact Evaluation:

- Enhance capabilities for evaluating the impact of internet use on various aspects of life, including age, education, and gender.
- Develop methods to ensure inclusivity in internet access and usage.

#### 3. Policy and Intervention:

- Develop effective policies and interventions to improve digital inclusivity and address digital skill difficulties.
- Strengthen the ability to conduct evaluations and assess the effectiveness of interventions.

#### 4. Research Methodology:

- Enhance research methodologies to conduct systematic literature reviews and perform valid comparative studies.
- Develop skills to employ various research methods, such as surveys and interviews.

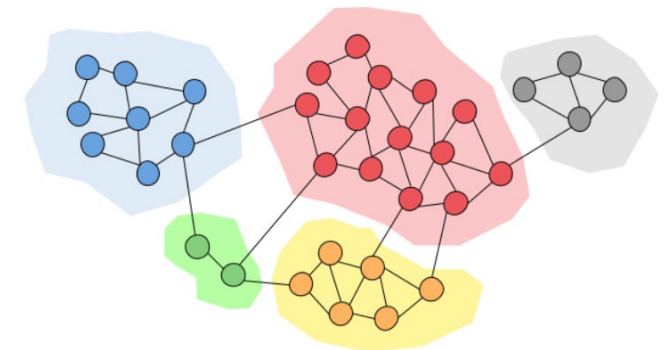
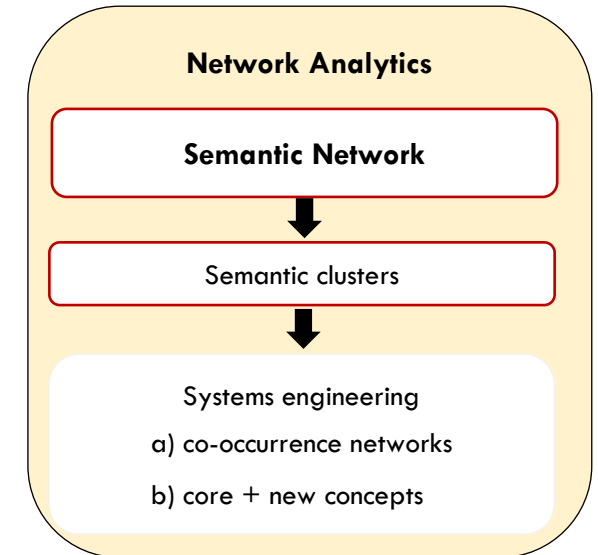
# Network Analytics

- **Semantic network**

- **Nodes:** most frequent bigrams
- **Links:** if bigrams appear within proximity of size 10
  - **Links weight:** weighted frequency

- **Semantic clusters**

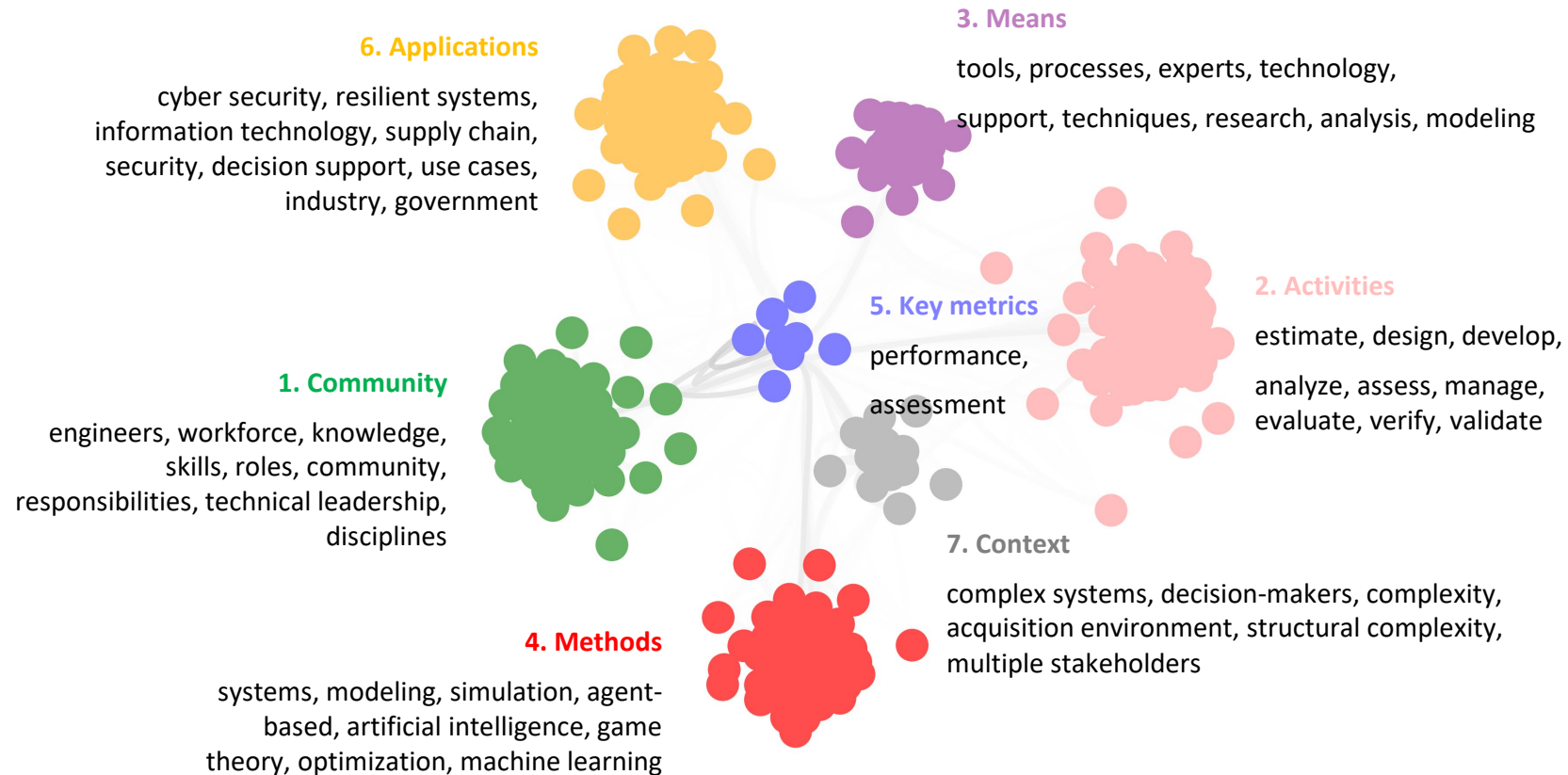
- Community detection
- Algorithm: Louvain community detection
- It starts with small communities and iterative calculates the modularity by adding and removing nodes to a community.
- Groups nodes with higher modularity (+ dense clusters)



# Semantic Network



## 7 Clusters



# Some Report Issues

- 2022 National Security Strategy 10.2022 – Password Protected
- Others: 10 reports so far not “keyword” relevant

## Verification (not in this project)

- Human-Based Summarization

# Reports with none or very limited Output

- DAU ACQ 0030 Objectives - Overview of Acquisition Ethics - 2 hours
- DAU ACQ 1010 Objectives - Fundamentals of Systems Acquisition Management - 13 hours
- DAU ACQ 0040 Objectives - Ethics Training - 1 hour
- DAU ACQ 1010 Objectives - Fundamentals of Systems Acquisition Management
- IDA 2019 - Importance of Data Analytics in Public and Private Sectors - NS D-10721
- IDA 2019 - Third Applying Data Analytics Forum D-10950
- NSF 2022 - SnT Public Perceptions Awareness Info Sources
- OPM 2022 Memo - Federal Workforce Priorities Report
- Phillips 2018 - Instruction Direction and Correction AWF - Defense ATnL
- Vincent 2021 - Digital DNA - in AF Chief SW Office Big Questions for Next CSO – Nextgov
- Wing 2019 - The Data Life Cycle - Harvard Data Science Review
- Baskin 2022 - 5 traits of the workforce of the future \_ MIT Sloan (No keywords)
- DeLong 2015 HBR - When Learning at Work Becomes Overwhelming (No keywords)
- Eastwood 2022 - New book explores how AI really changes the way we work \_ MIT Sloan (No keywords)
- Garamone 2022 DoD News - National Security Strategy (No keywords)
- Werber et al. 2019 - AWF Business Acumen and Training - RAND RR-2825-OSD (No Keywords)
- **World Economic Forum 2023 - The Future of Jobs**