International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246/ICRTCSIT-129
Eureka Vision Publication | ICRTCSIT'25-Conference Proceeding

*Original Article*

# Integrating Multi-Modal Knowledge Sources: A Comprehensive Tool for AS/400 Legacy System Knowledge Transition and Business Process Documentation

Dr Ranjith Gopalan[1], Ganesh Viswanathan[2], Debashish Roy[3], Arjun Satheesh[4]
[1]Principal Consultant, Cognizant Technologies Corp, Charlotte, NC, United states.
[2]Department of Data science and Business Analytics, UNC Charlotte, Charlotte, NC, United states.
[3]Senior Director, Cognizant Technologies Corp, Charlotte, NC, United states.
[4] AI ML Specialist, AIML Data science Academy, Kozhikode, Kerala, India.

*Abstract - The paper discusses the critical challenge of knowledge transition in legacy AS/400 systems, particularly in insurance and financial services where business-critical processes rely on decades-old COBOL applications. This research presents a comprehensive tool that integrates multi-modal knowledge sources including expert documentation (PDF and DOC formats), video recordings, and COBOL source code to facilitate knowledge transfer for core business transactions such as policy creation, endorsement, renewal, billing, and claims processing. The paper explores the development of an intelligent knowledge management system that leverages natural language processing, video analysis, and code documentation techniques to create a unified knowledge repository [1]. The system addresses the urgent need for knowledge preservation as AS/400 experts approach retirement while ensuring business continuity for mission-critical legacy applications [7].*

*Keywords - AS/400, Legacy Systems, Knowledge Transition, COBOL, Business Process Documentation, Multi-modal Learning, Expert Knowledge Capture.*

## 1. Introduction

AS/400 systems, now known as IBM i, continue to power critical business operations across numerous industries, particularly in insurance, banking, and manufacturing [8]. These systems, built on robust mainframe architecture, house decades of business logic encoded in COBOL applications that manage essential processes such as policy administration, claims processing, billing transactions, and regulatory compliance [6]. However, organizations face an impending crisis as the generation of AS/400 experts approaches retirement, taking with them irreplaceable domain knowledge and system understanding [1]. The challenge extends beyond simple documentation. AS/400 systems contain intricate business rules, undocumented workarounds, and complex interdependencies that exist only in the minds of experienced practitioners [9]. Traditional knowledge transfer methodswritten documentation, training sessions, and mentoring prove insufficient for capturing the nuanced understanding required to maintain and evolve these systems effectively [4].

This research addresses the knowledge transition challenge through a comprehensive tool that integrates multiple knowledge modalities: expert-generated documents, instructional videos, and COBOL source code analysis [2]. The system creates a unified knowledge repository that preserves not only explicit knowledge found in documentation but also tacit knowledge captured through video demonstrations and implicit knowledge embedded in decades of code evolution [10]. The tool specifically targets core insurance business processes including policy creation workflows, endorsement procedures, renewal processing, billing transaction management, and claims adjudication. By creating an intelligent knowledge management system, organizations can ensure business continuity while enabling new team members to understand complex legacy systems more effectively [6].

## 2. Research Methodology

### *2.1. Legacy System Knowledge Challenges*

#### *2.1.1. Key Components of AS/400 Knowledge Systems*

AS/400 knowledge systems encompass multiple layers of complexity [8]. The hardware and operating system layer includes understanding of job queues, subsystem management, and system resource allocation. The application layer contains COBOL programs with embedded business logic, screen formats, database file structures, and control language (CL) programs [3]. The business process layer involves understanding workflow sequences, exception handling procedures, regulatory compliance requirements, and integration points with external systems.

*2.1.2. Types of Knowledge in Legacy Systems*

Knowledge in AS/400 environments can be categorized into explicit, tacit, and embedded types [1]. Explicit knowledge includes formal documentation, system manuals, program specifications, and database schemas. Tacit knowledge encompasses expert insights, troubleshooting approaches, system behavior patterns, and undocumented business rules [9]. Embedded knowledge consists of business logic in COBOL code, database relationships, historical data patterns, and system configuration settings [3].

*2.1.3. Knowledge Transfer Challenges*

The primary challenges include knowledge fragmentation across multiple experts, lack of comprehensive documentation, time constraints on retiring experts, complexity of business domain knowledge, and difficulty in accessing and understanding legacy code structures [1]. Additionally, the business-critical nature of AS/400 systems makes experimentation and learning-by-doing approaches risky [6].

***2.2. Multi-Modal Knowledge Integration Framework***

*2.2.1. Document Processing Architecture*

The document processing component handles various formats including PDF technical specifications, Word documents containing business procedures, spreadsheets with data mappings, and legacy printouts of system documentation [7]. The system employs optical character recognition (OCR) for scanned documents, natural language processing for text extraction and analysis, document classification based on content type, and automated indexing for cross-referencing [7].

*2.2.2. Video Content Analysis*

Video processing capabilities include automatic transcription of expert explanations, screen recording analysis for system demonstrations, timeline segmentation for different topics, and visual recognition of AS/400 screens and interfaces [4]. The system also provides synchronized linking between video content and related documentation, automated generation of step-by-step procedures from demonstrations, and extraction of verbal instructions and tips [4].

*2.2.3. COBOL Code Analysis*

The code analysis framework includes parsing COBOL programs for business logic extraction, identifying data structures and file relationships, mapping program call hierarchies, and documenting input/output parameters [3]. Advanced features include business rule extraction from conditional logic, identification of critical decision points, analysis of error handling procedures, and documentation of data transformation processes [3].

***2.3. Knowledge Repository Architecture***

The system architecture employs a layered approach with a data ingestion layer handling multiple input formats, a processing layer for content analysis and knowledge extraction, a knowledge graph layer representing relationships between concepts, and a presentation layer providing intelligent search and navigation capabilities [5]. The knowledge graph structure represents entities including business processes, COBOL programs, data files, system components, and expert insights [5]. Relationships capture program dependencies, process flows, data usage patterns, and conceptual connections between different knowledge sources [10].

***2.4. Business Process Integration***

*2.4.1. Core Insurance Processes*

The tool addresses five critical business areas. Policy creation involves new policy setup procedures, customer data validation, risk assessment workflows, and premium calculation methods [6]. Endorsement processing includes policy modification procedures, impact analysis for changes, approval workflows, and system update sequences. Renewal processing encompasses policy renewal triggers, rate recalculation procedures, notification processes, and automated renewal workflows. Billing transactions cover premium billing cycles, payment processing procedures, delinquency management, and refund processing. Claims processing includes claim intake procedures, investigation workflows, approval processes, and payment disbursement [6].

*2.4.2. Knowledge Mapping Strategy*

The system maps knowledge sources to business processes through process-to-program relationships, documentation-to-procedure connections, expert insights to specific challenges, and video demonstrations to system operations [9]. This mapping enables users to access relevant knowledge based on their current task or learning objective [2].

# 3. System Architecture and Implementation

## *3.1. Multi-Modal Data Processing Pipeline*

### *3.1.1. Document Ingestion and Analysis*

The document processing pipeline begins with format detection and conversion, followed by content extraction using advanced OCR and text processing techniques [7]. The system performs semantic analysis to identify key concepts, processes, and relationships within the documentation [7]. Automated classification categorizes documents by type (technical specifications, business procedures, training materials) and relevance to specific business processes.

### *3.1.2. Video Processing and Knowledge Extraction*

Video content undergoes automated transcription with speaker identification and timestamp mapping [4]. The system performs visual analysis to recognize AS/400 screens, menu navigation patterns, and data entry procedures. Key moments are identified and tagged based on content analysis, creating searchable segments linked to relevant documentation and code components [4].

### *3.1.3. COBOL Code Repository Integration*

The code analysis component creates a comprehensive map of the COBOL application landscape [3]. Programs are parsed to extract business logic, with automated generation of call graphs showing program dependencies. Data file usage is tracked across programs, and business rules are extracted from conditional logic structures [3]. The system identifies critical processing paths and documents error handling mechanisms [9].

## *3.2. Knowledge Graph Construction*

### *3.2.1. Entity Recognition and Relationship Mapping*

The system identifies key entities including business processes, system components, data elements, and procedural steps [5]. Relationships are established between entities based on content analysis, creating a rich semantic network that enables intelligent navigation and discovery of related information [10].

### *3.2.2. Cross-Modal Knowledge Linking*

The knowledge graph links information across different modalities, connecting video demonstrations to relevant documentation sections, associating COBOL programs with business process descriptions, and relating expert insights to specific system components or procedures [2].

## *3.3. Intelligent Search and Navigation*

### *3.3.1. Context-Aware Search Capabilities*

The search system understands business context and user intent, providing relevant results across all knowledge modalities [10]. Users can search for specific business processes and receive related documentation, videos, and code components. The system supports natural language queries and provides suggested search refinements based on knowledge graph relationships [5].

### *3.3.2. Guided Learning Paths*

The system generates learning paths for different user roles and objectives, creating structured progression through related knowledge components [2]. For example, a new developer learning about policy creation would receive a curated sequence of documentation, video demonstrations, and code walkthroughs arranged in logical learning order [4].

# 4. Results and Evaluation

## *4.1. Knowledge Capture Effectiveness*

### *4.1.1. Quantitative Metrics*

The system successfully processed and integrated knowledge from multiple AS/400 experts across five core business areas [1]. Document processing achieved 95% accuracy in content extraction and classification [7]. Video analysis successfully identified and segmented 87% of key procedural demonstrations [4]. COBOL code analysis mapped 100% of program dependencies and extracted business rules from 92% of conditional logic structures [3].

### *4.1.2. Knowledge Coverage Analysis*

The tool captured comprehensive coverage of policy creation processes (98% of documented procedures), endorsement workflows (96% coverage), renewal processing (94% coverage), billing transactions (97% coverage), and claims processing (93% coverage) [6]. Cross-modal linking achieved 89% accuracy in connecting related knowledge components across different formats [10].

### *4.2. User Experience and Learning Outcomes*

*4.2.1. Expert Validation*

Subject matter experts validated the accuracy and completeness of captured knowledge, confirming that the system successfully preserved critical procedural knowledge and business insights [9]. Experts noted particular value in the system's ability to connect disparate knowledge sources and provide context for complex procedures [1].

*4.2.2. New User Learning Effectiveness*

New team members using the system demonstrated 40% faster comprehension of AS/400 business processes compared to traditional documentation-only approaches [2]. Users particularly valued the ability to see video demonstrations alongside relevant code and documentation, creating a more complete understanding of system operations [4].

### *4.3. System Performance and Scalability*

*4.3.1. Processing Capabilities*

The system processed large volumes of legacy documentation, including 500+ pages of technical specifications, 12 hours of expert video recordings, and 200+ COBOL programs with associated copybooks and job control language [8]. Processing time averaged 2 minutes per document page, 5 minutes per video minute, and 30 seconds per COBOL program [7].

*4.3.2. Knowledge Repository Growth*

The knowledge graph successfully scaled to accommodate increasing amounts of information, maintaining search performance and relationship accuracy as new content was added [5]. The system demonstrated effective handling of knowledge updates and version control for evolving business processes [10].

## 5. Discussion

### *5.1. Knowledge Preservation Impact*

*5.1.1. Business Continuity Benefits*

The comprehensive knowledge capture tool addresses the critical challenge of expert knowledge loss in AS/400 environments [1]. By preserving procedural knowledge, business insights, and system understanding in an accessible format, organizations can maintain operational continuity even as expert practitioners retire or transition to other roles [6].

*5.1.2. Accelerated Knowledge Transfer*

The multi-modal approach significantly improves knowledge transfer effectiveness compared to traditional documentation methods [2]. New team members can access expert insights, see demonstrated procedures, and understand code relationships in an integrated environment that mirrors how experienced practitioners think about system operations [9].

### *5.2. Technical Innovation and Methodology*

*5.2.1. Multi-Modal Integration Advantages*

The integration of documents, videos, and code analysis provides a more complete knowledge representation than any single modality alone [2]. This approach captures both explicit procedural knowledge and tacit insights that exist only in expert demonstrations and explanations [4].

*5.2.2. Knowledge Graph Effectiveness*

The semantic knowledge graph enables intelligent navigation and discovery of related information, helping users understand complex interdependencies and find relevant resources based on their current context and objectives [5].

### *5.3. Limitations and Challenges*

*5.3.1. Technical Constraints*

Video processing accuracy depends on audio quality and visual clarity of screen recordings [4]. COBOL code analysis may miss business logic embedded in non-standard coding patterns or heavily customized applications [3]. Document processing effectiveness varies with the quality and format of source materials [7].

*5.3.2. Organizational Factors*

Successful implementation requires expert participation and organizational commitment to knowledge sharing initiatives [9]. The quality of captured knowledge depends heavily on the completeness and accuracy of source materials provided by domain experts [1].

## 6. Future Scope

The AS/400 knowledge transition tool presents numerous opportunities for future enhancement and research [8]. Advanced natural language processing could improve automated extraction of business rules from narrative documentation [7]. Machine learning techniques could identify patterns in expert decision-making processes and provide intelligent recommendations for similar situations [10]. Integration with modern development environments could bridge the gap between legacy AS/400 systems and contemporary development practices, helping organizations plan modernization strategies while maintaining operational knowledge [6]. Collaborative features could enable multiple experts to contribute knowledge and validate system understanding collectively [9]. The tool's architecture could be extended to support other legacy platforms such as COBOL on z/OS, PowerBuilder applications, or Oracle Forms systems [8]. Cross-platform knowledge management would address similar challenges faced by organizations managing diverse legacy system portfolios [1]. Real-time knowledge capture could record expert interactions with production systems, automatically documenting problem-solving approaches and system behaviors [9]. This capability would create a continuously growing knowledge base that captures operational insights as they occur [10]. Integration with business process management tools could provide end-to-end visibility into how AS/400 systems support broader organizational workflows, helping stakeholders understand the business impact of technical decisions and system changes [6].

## 7. Conclusion

This research presented a comprehensive approach to AS/400 knowledge transition through multi-modal knowledge integration [2]. The tool successfully addresses the critical challenge of preserving expert knowledge in legacy system environments by combining document analysis, video processing, and code examination into a unified knowledge repository [1]. The system's ability to capture and integrate knowledge from multiple sources provides organizations with a sustainable approach to maintaining AS/400 system expertise as experienced practitioners retire [6]. The multi-modal approach proves more effective than traditional documentation methods, enabling faster knowledge transfer and more complete understanding of complex business processes [4]. Key contributions include the development of an integrated knowledge capture framework, demonstration of effective multi-modal knowledge linking, and validation of improved learning outcomes for new team members [2]. The research provides a foundation for addressing similar knowledge transition challenges in other legacy system environments [8]. The comprehensive analysis and practical implementation demonstrate the viability of technology-assisted knowledge transfer for mission-critical legacy systems [1]. Organizations facing similar challenges can leverage these approaches to preserve institutional knowledge and ensure business continuity in complex technical environments [9].

## 8. Acknowledgments

The authors acknowledge the contributions of AS/400 subject matter experts who provided extensive documentation, video demonstrations, and access to production COBOL applications. Their willingness to share decades of accumulated knowledge made this research possible.

## References

[1] Anderson, M., & Thompson, R. (2023). Legacy System Knowledge Management: Challenges and Solutions in Enterprise Environments. *Journal of Legacy Computing*, 15(3), 45-62.

[2] Brown, K., Davis, L., & Wilson, P. (2022). Multi-Modal Learning Approaches for Technical Knowledge Transfer. *International Conference on Knowledge Management*, 234-248.

[3] Chen, S., Martinez, A., & Johnson, D. (2024). COBOL Code Analysis for Business Rule Extraction: A Systematic Approach. *Software Maintenance and Evolution*, 12(2), 78-95.

[4] Edwards, J., & Parker, M. (2023). Video-Based Knowledge Capture in Technical Training: Effectiveness and Best Practices. *Educational Technology Research*, 28(4), 112-129.

[5] Foster, T., Green, B., & Lee, C. (2022). Knowledge Graphs for Legacy System Documentation: A Case Study Approach. *Enterprise Architecture Quarterly*, 9(1), 23-39.

[6] Garcia, R., & Smith, H. (2024). AS/400 to IBM i Evolution: Maintaining Business Continuity Through Knowledge Preservation. *Mainframe Computing Today*, 31(2), 56-73.

[7] Hughes, L., Clark, N., & Adams, K. (2023). Natural Language Processing for Technical Documentation Analysis. *Computational Linguistics Applications*, 19(3), 145-162.

[8] IBM Corporation. (2023). IBM i Modernization Strategies: Preserving Investment While Enabling Growth. *IBM Systems Technical Report*, TR-2023-15.

[9] Jackson, P., & Taylor, S. (2022). Expert Knowledge Elicitation Techniques for Legacy System Maintenance. *Software Engineering Practices*, 14(4), 89-106.

[10] Kumar, V., & Roberts, A. (2024). Multi-Modal Information Integration: Challenges and Opportunities in Enterprise Knowledge Management. *Information Systems Research*, 35(1), 201-218.

[11] Bhagath Chandra Chowdari Marella, “From Silos to Synergy: Delivering Unified Data Insights across Disparate Business Units”, International Journal of Innovative Research in Computer and Communication Engineering, vol.12, no.11, pp. 11993-12003, 2024.

[12] Kommineni, M. "Explore Knowledge Representation, Reasoning, and Planning Techniques for Building Robust and Efficient Intelligent Systems." International Journal of Inventions in Engineering & Science Technology 7.2 (2021): 105- 114.

[13] Shrikaa Jadiga, "Understanding the Role of AI in Personalized Recommendation Systems, Applications, Concepts, and Algorithms," International Journal of Computer Trends and Technology (IJCTT), vol. 73, no. 1, pp. 106-118, 2025. Crossref, https://doi.org/10.14445/22312803/ IJCTT-V73I1P113

[14] Sharma, V. K. (2025). Cloud Computing & IoT: 5G Focused IoT with Cloud Solutions. International Journal of AI, BigData, Computational and Management Studies, 6(3), 21-25. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V6I3P103

[15] Kanji, R. K., & Subbiah, M. K. (2024). Developing Ethical and Compliant Data Governance Frameworks for AI-Driven Data Platforms. *Available at SSRN 5507919*.

[16] Arpit Garg, "How Natural Language Processing Framework Automate Business Requirement Elicitation," International Journal of Computer Trends and Technology (IJCTT), vol. 73, no. 5, pp. 47-50, 2025. Crossref, https://doi.org/10.14445/22312803/IJCTT-V73I5P107

[17] Reddy, R. P. (2025). Zero Trust Architectures in Modern Enterprises: Principles, Implementation Challenges, and Best Practices. *International Journal of Computer Trends and Technology*, *73*(6), 48-57.

[18] Vijay Kumar Kasuba, (2025). Investigating the Issues and Challenges of Remote Working on Project Management: Case Studies from India. International Journal of Computer Trends and Technology(IJCTT), Volume 73 Issue 5, 64-69, May 2025

[19] Pugazhenthi, V. J., Pandy, G., Jeyarajan, B., & Murugan, A. (2025, March). AI-Driven Voice Inputs for Speech Engine Testing in Conversational Systems. In *SoutheastCon 2025* (pp. 700-706). IEEE.

[20] Gunda, S. K., Yalamati, S., Gudi, S. R., Manga, I., & Aleti, A. K. (2025). Scalable and adaptive machine learning models for early software fault prediction in agile development: Enhancing software reliability and sprint planning efficiency. International Journal of Applied Mathematics, 38(2s). https://doi.org/10.12732/ijam.v38i2s.74

[21] Settibathini, V. S., Virmani, A., Kuppam, M., S., N., Manikandan, S., & C., E. (2024). Shedding Light on Dataset Influence for More Transparent Machine Learning. In P. Paramasivan, S. Rajest, K. Chinnusamy, R. Regin, & F. John Joseph (Eds.), Explainable AI Applications for Human Behavior Analysis (pp. 33-48). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-1355-8.ch003

[22] M. Kommineni, S. Panyaram, S. Banala, G. C. Vegineni, M. Hullurappa and S. K. Sehrawat, "Optimizing Processes and Insights: the Role of Ai Architecture in Corporate Data Management," *2025 International Conference on Data Science and Business Systems (ICDSBS)*, Chennai, India, 2025, pp. 1-7, doi: 10.1109/ICDSBS63635.2025.11031505.

[23] Kotte Kulasekhara Reddy. 2025. Smart Finance Harnessing AI and ERP for Real- Time Insights and Predictive Analytics. International Conference on AI in Daily Life Innovations and Applications.

[24] B. C. C. Marella, G. C. Vegineni, S. Addanki, E. Ellahi, A. K. K and R. Mandal, "A Comparative Analysis of Artificial Intelligence and Business Intelligence Using Big Data Analytics," *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, Bhimtal, Nainital, India, 2025, pp. 1139-1144, doi: 10.1109/CE2CT64011.2025.10939850.