https://journal.ypidathu.or.id/index.php/humaniora
P - ISSN: 3062-7540
E - ISSN: 3048-3492

Digital Philology and Manuscript Sustainability: A Semantic Annotation Model for Classical Arabic Texts

Surip Stanislaus[1], Mona Abdallah[2]
[1]Universitas Katolik Santo Thomas Medan, Indonesia
[2]Alexandria University, Egypt

ABSTRACT

Background. The sustainability of classical Arabic manuscripts is often confined to digitization efforts that focus solely on image preservation and limited text markup. These approaches do not fully address the semantic richness and epistemological structure inherent in Islamic intellectual heritage.

Purpose. This study aims to develop a semantic annotation model tailored for classical Arabic texts to support digital philology and enrich manuscript sustainability through machine-readable and concept-linked interpretations.

Method. Using a developmental qualitative research design, three classical manuscripts were annotated semantically using a custom-built model based on RDF and Islamic ontology. The model was evaluated by domain experts in philology and computational linguistics, focusing on four criteria: semantic accuracy, contextual relevance, interoperability, and usability.

Results. The model achieved annotation accuracy above 91% across all manuscripts. Experts rated semantic precision (4.7/5) and contextual relevance (4.6/5) as its strongest aspects. The system successfully mapped technical terms and logical concepts within classical texts and linked them across manuscripts. A case study demonstrated the model's effectiveness in identifying relationships between epistemological terms and enabling thematic exploration.

Conclusion. This semantic annotation model advances digital philology by enabling structured, concept-based analysis of classical Arabic texts. It bridges computational methods with Islamic textual traditions and opens new pathways for collaborative, sustainable, and meaningful engagement with manuscript heritage.

KEYWORDS

Classical Arabic Manuscripts, Digital Philology, Semantic Annotation

Citation: Stanislaus, S., & Abdallah, M. (2025). Digital Philology and Manuscript Sustainability: A Semantic Annotation Model for Classical Arabic Texts. *Journal of Humanities Research Sustainability, 2*(3), 135–144.
https://doi.org10.70177/jhrs.v2i3.2304

Correspondence:

Surip Stanislaus,
suripofmcap@yahoo.com

Received: June 21, 2025

Accepted: July 2, 2025

Published: July 21, 2025

INTRODUCTION

Philology is a discipline that plays an important role in the preservation of the intellectual heritage of mankind, including classical Arabic texts that hold treasures of Islamic science, literature, law, and philosophy (Goto, 2024). Traditional philological activities focus on transcription, verification of manuscript variants, and the preparation of critical editions that require high precision and long time. This effort is the foundation in ensuring the authenticity and sustainability of the scientific values contained in these manuscripts (Keersmaekers, 2020).

The development of digital technology in the last two decades has opened up new horizons for philological practice, particularly in the context of the preservation and dissemination of ancient manuscripts (Al-Homed dkk., 2023). Digital philology emerged as an interdisciplinary field that combines philological methodologies with modern computing capabilities. The use of digital imagery, OCR (Optical Character Recognition) software, and online databases allows for wider access to manuscripts while speeding up the textual review process (ABBOU, 2025).

Classical Arabic texts present unique challenges due to their rich morphology, complex syntactic structure, and use of diverse terminological and rhetorical terms (Salah dkk., 2024). This uniqueness demands a more sophisticated technological approach compared to Latin or Greek texts. Digitizing text alone is not enough; a semantic annotation system is needed that is able to recognize, classify, and connect meanings contextually within the framework of classical Islamic science (Osman, 2020).

The theory of semantic annotation in computational linguistics states that the addition of semantic metadata to text allows for automatic processing of information by machines without losing its original context (McGillivray dkk., 2022). This model creates a data structure that is both machine and human-readable, making it particularly relevant to classical Arabic manuscripts that are rich in meaning and multi-layered in interpretation. In practice, semantic annotations highlight not only literal meanings, but also relationships between entities, concepts, and terms inherent in scientific discourse (Zouaoui & Rezeg, 2021).

Efforts towards the sustainability of classical Arabic texts depend not only on physical preservation or visual digitization, but also on the ability to interpret and represent those meanings semantically (Bos, 2022). Digital philology equipped with a semantic annotation model offers a new way to revive texts, not just as historical artifacts, but as a source of living knowledge that can be accessed across generations and across disciplines (Hassan, 2025).

Digitization of classical Arabic manuscripts has been widely carried out by libraries and Islamic studies centers, but only to scanning and converting visuals into digital form (Bensattalah dkk., 2023). This process does save physical forms and open access, but it has not touched on the aspects of meaning and the structure of knowledge contained in it. Digital manuscripts that are not semantically annotated remain difficult to make optimal use by cross-disciplinary researchers (Gruendler dkk., 2020) .

The annotation system that exists today is mostly developed for Indo-European languages and does not take into account the linguistic structures typical of classical Arabic such as patterns of derivation (i'rab), tarkib idhafah, or epistemic terms in Islamic scientific discourse (Taye dkk., 2023). As a result, classical Arabic texts are still read manually and require long and complex interpretations. The lack of a semantic annotation system that corresponds to the linguistic and scientific context of classical Arabic creates a huge gap between the potential of digital philology and its actual practice (Aliyev, 2024).

Manuscript preservation efforts often stop at the stage of physical conservation and visual digitization without paying attention to the sustainability of meaning and ease of exploration of content. Without a semantic annotation mechanism, academic interaction with digital manuscripts is still very limited. The search tool still relies on literal keywords, not the relationship of meaning or structure of concepts (Bouziane dkk., 2020).

The Linked Open Data Theory provides a strategic approach to connecting information between entities in text through URI-based semantic identification (Uniform Resource Identifier) (Muhammed dkk., 2024). This model allows the relationships between terms, topics, figures, and historical contexts in classical texts to be represented digitally and openly. In the context of Arabic

philology, this approach can answer the limitations of conventional search systems and open up opportunities for data collaboration across global manuscript collections (Elayeb, 2019).

Filling in the gaps in the semantic processing of classical Arabic texts is essential to expand access to knowledge of the Islamic heritage (Lima dkk., 2025). Classical manuscripts contain layers of meaning that are intertwined between language, logic, theology, and law. This meaning cannot be represented only through digital texts or conventional cataloging. An annotation system is needed that is able to capture the relationship between concepts and meanings in the context of classical science (Madi dkk., 2024).

The development of semantic annotation models specific to classical Arabic texts is a crucial step in building a more contextual and sustainable future of digital philology. This system allows for concept-based search, integration between works, and visualization of networks of thought contained in manuscripts. With it, researchers not only read texts, but can trace scientific relationships between manuscripts, topics, and even thinkers.

The Hermeneutics of the Digital Text theory states that digital interaction with classical texts is not only reproductive but also interpretive (Khan dkk., 2022). Technology enables the exploration of layers of meaning that were previously difficult to reach in a linear manner. In this context, semantic annotations are not just technical aids, but rather a bridge of interpretation between ancient texts and contemporary ways of thinking. The implementation of this theory in classical Arabic manuscripts is a strategic contribution to strengthening the intellectual sustainability of Islamic heritage manuscripts (Farina dkk., 2025).

RESEARCH METHODOLOGY

This study uses an exploratory qualitative approach with a developmental study design to design and test a semantic annotation model for classical Arabic texts (Liao & Zhao, 2019). The main focus of the research is to build an annotation system capable of recognizing linguistic structure and conceptual meaning in classical Arabic manuscripts, as well as testing its effectiveness in the context of digital philology. This design was chosen to allow iteration of the model design based on empirical findings from text review and user feedback (Kiraz, 2023).

The research population consists of classical Arabic manuscripts in the fields of jurisprudence, interpretation, and logic that have been digitized by online manuscript providers such as Al-Maktabah al-Shamilah, King Saud University, and the World Digital Library (Crane dkk., 2024). Samples were taken purposively from three representative manuscripts that had a complex text structure, a glossary of technical terms, and a diversity of Arabic morphological forms. The selection was based on the availability of digital versions of the text and the diversity of scientific themes to test the flexibility of the annotation model (Fedeli, 2020).

The research instruments include a prototype of web-based semantic annotation software developed using TEI (Text Encoding Initiative) and RDF (Resource Description Framework) standards. Additional instruments in the form of model feasibility test sheets and linguistic annotation guidelines were developed based on Arabic morphological parameters, taxonomy of Islamic scientific topics, and conceptual metadata frameworks. The validity of the instrument was tested through expert tests from the fields of Arabic philology and computational linguistics (Jreis-Navarro, t.t.).

The research procedure involves four main stages, namely: (1) analysis of the structure of texts and terms in classical Arabic manuscripts, (2) development of semantic annotation prototypes with concept, relational, and keyword entries, (3) application of the model to selected manuscripts, and (4) evaluation of the effectiveness of the model based on the criteria of readability, semantic

accuracy, and data interoperability. The results of the evaluation are used to refine the model and produce semantic annotation implementation guidelines that can be replicated by researchers or other digital manuscript providers (Sawalha, 2019).

RESULT AND DISCUSSION

Table 1. Semantic Annotation Results of Three Classical Manuscripts

Manuscript Title	Total Segments	Annotated Segments	Accuracy (%)
Nihayat al-Sul	1500	1395	93.0
Tafsir al-Kashani	1200	1132	91.4
Sharh al-Aqidah	950	890	93.7

These data show that the annotation system developed has a high level of reliability for recognizing linguistic structures and concepts in classical manuscripts, albeit with different thematic variations and writing styles.

Three classical Arabic manuscripts used as samples were successfully semantically annotated using the prototype model developed. Each manuscript has a different number of text segments depending on the complexity of the structure and the length of the content. Nihayat al-Sul consists of 1500 segments, Tafsir al-Kashani 1200 segments, and Sharh al-Aqidah 950 segments.

The number of segments that were successfully annotated indicates the level of effectiveness of the machine's work and the compatibility of the text structure with the annotation system. The highest percentage of accuracy was obtained in the manuscript of Sharh al-Aqidah at 93.7%, followed by Nihayat al-Sul at 93%, and Tafsir al-Kashani at 91.4%. Accuracy is calculated based on the match of the results of machine annotations with the validation of Arabic philology experts.

The difference in accuracy between manuscripts is influenced by the type of discourse and the complexity of the terms used. Manuscripts with a predominance of descriptive language such as religious lectures are easier to annotate than interpretations that are rich in symbolic interpretations. The argumentative nature and cross-quotations in interpretation pose a greater challenge to the recognition of semantic patterns.

The success of annotations in more than 90% of segments shows that this model can work well on a wide range of genres of classical text. The use of a combination of semantic tags for entity figures, places, and technical terms successfully improves the accuracy of identification of key concepts in the text. This annotation system proved to be adaptive to the structure of i'rab and classical rhetorical styles.

The flexibility of the model is also supported by RDF's relational structure that allows for cross-segmentation and metadata correlation. This makes it easier to explore concepts and topics in the context of the verses, hadiths, or logic being explained. Data connectivity is an important factor in reviving the scientific structure in classical texts.

The evaluation of the feasibility of the model was carried out through a panel of experts consisting of three Arabic philologists, one computational linguist, and one annotation software developer. The assessment was carried out based on four main dimensions, namely semantic accuracy, ease of use, system interoperability, and contextual relevance to classical texts. Each dimension is rated on a scale of 1–5.

The results of the evaluation showed that the semantic precision dimension scored the highest with an average of 4.7. The contextual relevance aspect is also highly rated, signaling the model's

ability to understand the context of terms in classical narratives. The interoperability of the system scored 4.5, reflecting the model's ability to connect with a digital manuscript data management system.

The ease of use dimension scored the lowest among the four categories, namely 4.3. This shows the need to improve the user interface and data input process to make it more practical to use by non-programmers or academics from non-technical fields.

A high score on semantic accuracy proves that the model is able to effectively capture terminological meanings and relationships between entities. The use of the RDF framework and *the ontology-aware tagging* concept provides a deeper meaning representation structure than ordinary literal annotations. Experts state that the system is able to distinguish between terms that are morphologically identical but conceptually different.

Table 2. Expert Assessment of the Feasibility Dimensions of Semantic Annotation Models

Criteria	**Expert Score (out of 5)**
Semantic Accuracy	4.7
Usability	4.3
Interoperability	4.5
Contextual Relevance	4.6

The interoperability dimension is an important aspect because it shows that the model can be integrated with digital library platforms such as TEI, Al-Maktabah Syamilah, and open source-based databases. This positive assessment opens up opportunities for wider use in academic and non-academic environments.

The value on the usability dimension indicates the need for improvements in system visualization and documentation. Some experts suggest adding a drag-and-drop system or a more intuitive visual annotation interface to make it easier for non-technical annotators to work. The development of a GUI (Graphical User Interface) is an advanced priority of this system.

Annotations and expert evaluation scores indicate that the technical success of the system supports conceptual acceptance of the model. Manuscripts with a more orderly structure obtained more stable annotation results and were judged to be more contextual. This relationship reinforces the assumption that the power of the model depends on the interaction of linguistic structures and conceptual tags.

The annotation system successfully applied to classical texts displays a pattern of semantic relationships that can be visualized as a network of concepts. This connectivity facilitates the exploration of intertextual knowledge between works. The integration of RDF and conceptual metadata allows users to systematically search for the thoughts of characters, specific terms, or scientific themes.

This model serves a dual function, as a digital preservation tool and as a bridge of textual interpretation. The relationship of data in the system supports the formation of a living digital knowledge base that can be expanded by the scientific community. Thus, semantic annotation paves the way to a dynamic, cross-disciplinary collaborative philology.

A case study was conducted on one segment of the *manuscript of Nihayat al-Sul*, part of the discussion of the logic of the syllogism (*qiyas*). This segment consists of 12 paragraphs with various technical terms such as *muqaddimah*, *natijah*, and *qadiyah mujibah*. Previously, researchers

had difficulty navigating the repetition of terms in different contexts and the intersectional associations.

Semantic annotation models are used to mark logical terms and relate them to definitions and examples in other segments. The results showed that the system managed to identify seven of the eight key terms automatically and associate two of them with similar entries in other manuscripts. The navigation process becomes more efficient, and the understanding of context improves.

Annotators state that the system is able to distinguish the meaning of terms when used in comparison (analogy) versus in syllogistic reasoning. This capability is not possible by conventional text search systems. This case study shows the great potential of models in supporting cross-text thematic studies.

The application of the system to logical texts proves that semantic annotations can work not only in religious or historical terms, but also in the field of abstract thought. The conceptual structure of dense, repetitive classical logic is an appropriate testing ground for assessing the semantic sensitivity of the model. The positive results strengthen the technical validity of the system in high-complexity text processing.

Identifying the relationship between terms makes it easier to map the thinking of classical text writers. Researchers can quickly trace cross-references in a single manuscript or across manuscripts. This data connection is very important in building the foundation of classical text-based digital thematic studies.

The effectiveness of annotations in case studies also suggests that this system is feasible to be developed as a critical note writing aid or digital scientific edition. Researchers can add second-layer annotations, connecting the text to modern interpretations or contemporary philosophical contexts. With this, the system serves not only as a preserver, but also as a developer of knowledge.

The linkages between the success of technical annotations, expert validation, and case study results form a pattern of consistency that demonstrates the feasibility and effectiveness of the model. Quantitative and qualitative data mutually reinforce that RDF-based semantic annotation is able to dynamically revive meaning in classical texts. The system is able to capture the complexity of language and manage meaning adaptively.

The annotation model not only successfully reads the text, but also builds a network of meanings that facilitates conceptual understanding. The relationship between terms, figures, and concepts is the basis for creating a digital system of exploration of classical Islamic knowledge. This brings the practice of philology in a more interactive and participatory direction.

This combination of technical and conceptual results marks the great potential of the development of semantics-based digital philology as a new avenue in the preservation and study of Arabic manuscripts. The results of this research open up a space for innovation in building a philological platform that is able to read, understand, and reconstruct the heritage of classical Islamic science in a sustainable and time-long manner.

The semantic annotation model developed in this study showed high effectiveness in identifying and classifying text segments from classical Arabic manuscripts. The annotation accuracy rate reached more than 91% for all manuscripts, with the highest scores on faith-themed texts and logic. The system manages to recognize technical terms, syntactic structures, and meaning contexts accurately and relevantly.

Evaluations from experts gave high scores mainly on the dimensions of semantic accuracy and contextual relevance. This score indicates that the system is not only technically effective, but also conceptually and methodologically valid from an academic perspective. These results are

supported by case studies that show significant improvements in navigation and understanding of texts by researchers.

The application of the system in case studies also reveals the ability of semantic annotations to construct a network of meanings that facilitates cross-segment and cross-script exploration. This ability is key in reviving classical texts not only as objects of study, but as an active knowledge base in a digital environment.

This research expands the approach of digital philology by adding layers of RDF-based semantics and classical Islamic ontology that have not been touched much in previous studies. Previous research generally stopped at visual digitization or simple TEI-based markup annotations. The main difference lies in the focus on the structure of meaning, not just the representation of the text.

The results of this study are in harmony with the linked *data approach* developed in classical Latin or Greek texts, but show additional complexity due to the morphological and syntactic properties typical of the Arabic language. This research proves that the semantic approach that has been successful in the West can be adapted to the context of classical Islam with structural and conceptual modifications.

Its main contribution lies in the incorporation of information technology with the distinctive principles of Arabic philology and semiotics. The application of this system proves that classical texts can be given digital meaning with precision without ignoring the complexity of the scientific tradition inherent in them.

The success of this annotation model is a sign that the preservation of manuscripts is no longer enough with the digitization of surfaces, but must enter the level of meaning that can be processed, traced, and redeveloped. Classical texts are not just historical sources, but maps of knowledge that can be revived through semantic technology.

These results also reflect a new need in the study of digital Islam to move from a descriptive approach to a conceptual and interactive approach. Researchers no longer just read texts, but navigate networks of meanings that enrich understanding and trigger reinterpretation.

This model is proof that digital philology is not only technical, but also epistemological. The transformation of form must be accompanied by a transformation of the way of thinking, so that the heritage of classical Islamic science can answer the challenges of the times and be relevant to the next generation of academics.

This semantic annotation model can be the basis for the development of an open and collaborative Arabic-language digital philology platform. Manuscript institutions can leverage it to enrich metadata, enhance scientific search, and support intellectual reconstruction across manuscripts. This implication is very important in the context of Islamic digital humanities.

The use of this system allows learning and research on classical texts to be carried out more efficiently, transparently, and accurately. Students and researchers can now search scientific concepts semantically, not just based on literal keywords. This paved the way for interdisciplinary, cross-era, and cross-regional studies.

In the long run, this system can strengthen the sustainability of manuscripts as a dynamic source of knowledge. The textual heritage of Islam is not only visually preserved, but revived in a format that supports global collaboration, digital learning, and ongoing scientific reconstruction.

The highly grammatically organized structure of classical Arabic allows semantic annotations to work effectively if supported by the right linguistic framework. The presence of i'rab, idhafah, and epistemic terminology patterns provides a stable foundation for contextual automatic reading.

The system is able to recognize the relationship of meaning because its structure has been codified in the tradition of nahwu and balaghah.

The use of RDF and the ontology-aware tagging approach makes the system recognize not only words, but also the relationships between concepts and entities contained in the text. This is what distinguishes semantic annotation from regular markup annotations. The system can understand that the term *qiyas* in logic has a hierarchical relationship and association of concepts with *muqaddimah* and *natijah*.

The involvement of experts in the validation stage provides valuable input to maintain a balance between technical structure and fidelity to philological meaning. The system was developed not only in the technology laboratory space, but through a collaborative process between computer science and Islamic sciences. This synergy is what produces stable, valid, and applicable results across manuscripts and disciplines.

Advanced development needs to focus on GUI-based user interfaces to make the system more accessible to non-technical researchers. Visualization of semantic networks in the form of graphs or mindmaps is also an important part to support the intuitive exploration of the structure of thought in manuscripts. This step will expand the reach of system users.

It is important to develop a broader and more open database of Islamic scientific ontology. Collaboration between universities, manuscript centers, and global Islamic institutions can form a standardized and interconnected annotation ecosystem. This system can be the foundation for Islamic scientific data linked projects on a global scale.

Future research may also evaluate the impact of using these systems in the context of teaching classical texts in an academic setting. Pedagogical experiments with semantic annotation systems can be a transformational step in bringing the younger generation closer to ancient texts through a more relevant and in-depth digital approach.

CONCLUSION

This study shows that the developed semantic annotation model is able to identify, classify, and associate meanings in classical Arabic texts with high accuracy and strong conceptual relevance. The most important finding of this study is its ability to systematically construct a network of meanings and concepts, thus allowing a re-reading of classical texts as living and traceable knowledge entities across manuscripts.

The main contribution of this research lies in a methodological approach that combines RDF frameworks, ontology-based semantic annotations, and classical philological principles to build a system of exploration of meaning in Arabic texts. This model presents a new form of digital philology that is not only documentary but also interpretive, with the potential to be developed as a collaborative platform for Islamic digital science.

The main limitation of this study lies in the limited number of manuscripts and the lack of a GUI-based visual user interface. Further research can be focused on the development of graph-based semantic visualizations, the expansion of the ontology of Islamic scientific topics, as well as the integration of systems into philology learning in the academic environment to be more adaptive to users from non-technical backgrounds.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.

REFERENCES

ABBOU, T. (2025). A DESCRIPTIVE STUDY OF THE PLATFORM OF ALGERIAN MANUSCRIPTS. *Asian &African Studies (13351257)*, *Query date: 2025-07-21 13:50:09*. https://www.sav.sk/journals/uploads/05301250aas2025-1_05_Abbou_kor3.pdf

Al-Homed, L., Jambi, K., & Al-Barhamtoshy, H. (2023). A deep learning approach for Arabic manuscripts classification. *Sensors*, *Query date: 2025-07-21 13:50:09*. https://www.mdpi.com/1424-8220/23/19/8133

Aliyev, C. (2024). An Overview of Al-Mutanabbi's Diwan Manuscripts: Catalog Information and Global Library Holdings. *Global Spectrum of Research and Humanities*, *Query date: 2025-07-21 13:50:09*. https://gsrh.net/index.php/home/article/view/35

Bensattalah, A., Chalal, R., & Nader, F. (2023). An Adaptive Semantic Annotation Tool for Teachers Based on Context-Aware and Internet of Things. *International Journal of Software …*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1142/S0218194023500341

Bos, J. (2022). A semantically annotated corpus of tombstone inscriptions. *International Journal of Digital Humanities*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1007/s42803-021-00039-y

Bouziane, A., Bouchiha, D., & ... (2020). Annotating Arabic texts with linked data. *2020 4th International …*, *Query date: 2025-07-21 13:50:09*. https://ieeexplore.ieee.org/abstract/document/9416543/

Crane, G., Tauber, J., Babeu, A., Cerrato, L., & ... (2024). The Sixth Generation of the Perseus Digital Library and a Workflow for Open Philology—DRAFT. *arXiv preprint arXiv …*, *Query date: 2025-07-21 13:50:09*. https://arxiv.org/abs/2411.10604

Elayeb, B. (2019). Arabic word sense disambiguation: A review. *Artificial Intelligence Review*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1007/S10462-018-9622-6

Farina, A., Marongiu, P., Bru, M., & ... (2025). When data meets the past: Data collection, sharing, and reuse in ancient world studies. *Open Information …*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1515/opis-2025-0014

Fedeli, A. (2020). The Qur'ānic Text from Manuscript to Digital Form: Metalinguistic Markup of Scribes and Editors. *Judaism, Christianity, and Islam–Tension …*, *Query date: 2025-07-21 13:50:09*. https://library.oapen.org/bitstream/handle/20.500.12657/41217/9783110634440.pdf?sequence=1&isAll#page=232

Goto, K. (2024). 11. Philology and Grammar. *Journal for the Study of the Old Testament*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1177/03090892241240277

Gruendler, B., Ginkel, J., Redwan, R., & ... (2020). An interim report on the editorial and analytical work of the AnonymClassic project. *medieval …*, *Query date: 2025-07-21 13:50:09*. https://refubium.fu-berlin.de/handle/fub188/33267

Hassan, R. (2025). AI Analysis of Islamic History Sources: Challenges and Opportunities. *Computing, Internet of Things and Data Analytics …*, *Query date: 2025-07-21 13:50:09*. https://books.google.com/books?hl=en&lr=&id=bRlxEQAAQBAJ&oi=fnd&pg=PA287&dq=semantic+annotation+classical+arabic+manuscripts+digital+philology&ots=4FeUJ0QtVj&sig=TM00YS0TOQyN7uy6C9m83_SX3aU

Jreis-Navarro, L. (t.t.). The last pre-modern Arab splendor over the Euro-African Strait: Renaissance" gap" and digital transition. *academia.edu*, *Query date: 2025-07-21 13:50:09*. https://www.academia.edu/download/84060471/PremodernArabSplendor_LailaJreis_AcademiaEdu.pdf

Keersmaekers, A. (2020). *A computational approach to the Greek papyri: Developing a corpus to study variation and change in the post-classical Greek complementation system*. *Query date: 2025-07-21 13:50:09*. https://lirias.kuleuven.be/retrieve/590983

Khan, A., Chiarcos, C., Declerck, T., Gifu, D., & ... (2022). When linguistics meets web technologies. Recent advances in modelling linguistic linked data. *Semantic …*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.3233/SW-222859

Kiraz, G. (2023). Towards a Syriac Semantic Web from the Perspective of 2020. *The Third Lung: New Trajectories in Syriac Studies …*, *Query date: 2025-07-21 13:50:09*. https://brill.com/downloadpdf/display/title/64228.pdf#page=290

Liao, X., & Zhao, Z. (2019). Unsupervised approaches for textual semantic annotation, a survey. *ACM Computing Surveys (CSUR)*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1145/3324473

Lima, B., Omar, N., Avansi, I., & Castro, L. de. (2025). Artificial Intelligence Applied to the Analysis of Biblical Scriptures: A Systematic Review. *Analytics*, *Query date: 2025-07-21 13:50:09*. https://www.mdpi.com/2813-2203/4/2/13

Madi, B., Atamni, N., Tsitrinovich, V., & ... (2024). Automated Dating of Medieval Manuscripts with a New Dataset. *... on Document Analysis …*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1007/978-3-031-70642-4_8

McGillivray, B., Kondakova, D., Burman, A., & ... (2022). A new corpus annotation framework for Latin diachronic lexical semantics. *... of Latin Linguistics*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1515/joll-2022-2007

Muhammed, M., Azab, S., Ali, N., & ... (2024). Arabic Ontology for Hadith texts-A survey. *The Egyptian Journal of …*, *Query date: 2025-07-21 13:50:09*. https://journals.ekb.eg/article_352397.html

Osman, A. (2020). *A methodological approach to utilize Egyptian colloquial Arabic as a source for ancient Egyptian linguistic analysis*. search.proquest.com. https://search.proquest.com/openview/be7074e73a4ad6ff705c747cc8b6e28e/1?pq-origsite=gscholar&cbl=2026366&diss=y

Salah, R., Mukred, M., Zakaria, L. binti, & ... (2024). A Machine Learning Approach for Named Entity Recognition in Classical Arabic Natural Language Processing. *KSII Transactions on …*, *Query date: 2025-07-21 13:50:09*. https://koreascience.kr/article/JAKO202433743226435.page

Sawalha, M. (2019). The Design and the Construction of the Traditional Arabic Lexicons Corpus (The TAL-Corpus). *Modern Applied Science*, *Query date: 2025-07-21 13:50:09*. https://www.researchgate.net/profile/Majdi-Sawalha/publication/330252539_The_Design_and_the_Construction_of_the_Traditional_Arabic_Lexicons_Corpus_The_TAL-Corpus/links/5c47782092851c22a3896daa/The-Design-and-the-Construction-of-the-Traditional-Arabic-Lexicons-Corpus-The-TAL-Corpus.pdf

Taye, M., Abulail, R., & Al-Oudat, M. (2023). An ontology learning framework for unstructured arabic text. *2023 7th International …*, *Query date: 2025-07-21 13:50:09*. https://ieeexplore.ieee.org/abstract/document/10391548/

Zouaoui, S., & Rezeg, K. (2021). A novel quranic search engine using an ontology-based semantic indexing. *Arabian Journal for Science and Engineering*, *Query date: 2025-07-21 13:50:09*. https://doi.org/10.1007/s13369-020-05082-5