Digital Humanities Overview in Lithuania: Heritage and Language Resources

Andrius UTKA[1], Giedrė VALŪNAITĖ-OLEŠKEVIČIENĖ[2], Anželika GAIDIENĖ[3]

[1]Institute of Digital Resources and Interdisciplinary Research (SITTI), Vytautas Magnus University, Kaunas, Lithuania
[2]Mykolas Romeris University, Vilnius, Lithuania
[3]Institute of the Lithuanian Language, Vilnius, Lithuania

andrius.utka@vdu.lt, gvalunaite@mruni.eu, anzelika.gaidiene@lki.lt

ORCID 0000-0001-5212-4310, ORCID 0000-0001-5688-2469, ORCID 0000-0001-8775-788X

Abstract: The paper provides an overview of the digital humanities landscape in Lithuania, highlighting key initiatives, strategies, and resources that contribute to the field's development. It begins by defining the broad and evolving scope of digital humanities, examining its interdisciplinary nature and the challenges in establishing clear boundaries of the field. Admitting a major role of digital humanities in preserving cultural and national heritage, the paper overviews key initiatives in Lithuania, prominent heritage digitization platforms, along with significant language resources. The paper also presents several digital humanities' use cases, including offensive language identification and digital archives of Lithuanian literature. The conclusion reflects on future trends, emphasizing the increasing role of artificial intelligence and the continued digitization of cultural heritage in advancing digital humanities in Lithuania.

Keywords: digital humanities, cultural heritage digitisation, Lithuanian language resources, heritage platforms

1. Introduction

1.1. Defining the Landscape of Digital Humanities

In the current world of technological advancement and the explosion of data, digital humanities (DH) play an important role in bridging the gap between traditional humanities disciplines and modern technologies. Although it is difficult to define the field of DH due to its vastness and interdisciplinary nature, some characteristics undoubtedly define this field: the usage of data and modern technology that serve the needs and address the issues of the humanities.

The landscape of DH has been overviewed in a number of studies by focusing both on the general landscape and typologies of the field (Svensson, 2010) and dwelling into particular DH landscapes of various countries. Burdick et al. (2016) extensively analyse the vast field of DH suggesting that the core of the discipline heavily relies on text-based

studies including literature, philosophy, history, etc. However, considering the new trends of merge between the digital and humanities in virtual environments and augmented reality, DH acquire new trajectories. Helmi (2021) proposes a broad approach presenting DH as interdisciplinary research with plethora of data and previously unexplored knowledge.

Furthermore, Davidson (2008) links the phases of development of DH to the evolution of the web, introducing the concepts of Humanities 1.0 and Humanities 2.0. The author observes that Humanities 2.0 excels by interactivity and technological augmentation compared to the data-based nature of Humanities 1.0. McPherson (2009) introduced a typology for the DH by distinguishing three typological classes: the computing humanities, blogging humanities, and multimodal humanities. The author characterizes the computing humanities as related to building tools, data, and infrastructure; the blogging humanities, which the author associates with networked media and the multimodal humanities are presented as embracing the features of both visual and aural media by focusing on the potential of multimodality. Both authors envision the developing nature and inclusive character of DH leaving space for constant development in the landscape of the field.

Following the perception of the lack of defined boundaries of DH Holm et al. (2015b) suggest five major research areas concerning DH:

- Digital collections, archiving and text encoding which is related to the aggregation of digital editions, digital corpora and existing data repositories;
- Adapting, augmenting and analysing electronic texts;
- Geospatial mapping technologies of complex spatial and temporal data dealing with the use geographic information systems to capture and analyse various forms of geographically related information;
- 'Big Data,' social computing, crowdsourcing, and networking connected to processing massive and complex data in DH field;
- 3D immersive visualisation environments pertinent to simulated augmented environments.

The attempts to characterize the landscape of DH reveal that the field is constantly developing and embracing a lot of changes related to the development of digital technologies and digital revolution transforming the humanities and broadening the boundaries of the research.

1.2. Digital Humanities in National Contexts

The role of DH in preserving and modernising cultural and national heritage is equally important. Successful DH projects attract not only the attention of contemporary society, but also disseminate cultural information and reinforce cultural identity. Several studies have explored the state of DH in various countries, such as. Toscano et al., (2020), who provide a comprehensive analysis of the situation in Spain, and Doran (2022) who offers an extensive report on Irish DH entities. Additionally, Hadalin (2014) presents a concise overview of DH in Slovenia, while Nygren et al. (2014) examine the status of DH in Sweden.

The **goal** of the current paper is to present an overview of the DH landscape in Lithuania, including important initiatives, strategies, and most important DH platforms and resources. As it is not possible to account for all different subfields of digital

humanities, we will focus on the intermixture of more prominent and influential heritage and language resources, tools, and projects in Lithuania.

2. Initiatives and Strategies for Promoting Digital Humanities in Lithuania

The Lithuanian government and other state institutions have prepared several programs for various language studies and to promote the progress of language technologies and digital humanities.

The Seimas of the Republic of Lithuania approved a document *The Guidelines for the Development of the Lithuanian Language in the Digital Environment and the Progress of Language Technologies for 2021–2027* (Guidelines, 2020). The Guidelines (2020) set out the essential tasks of Lithuanian language technologies. "They stipulate what should be done in Lithuania in the near future and in which directions to work: 1. To increase the competence of specialists working in the field of language technologies and to raise the level of society's ability to use the opportunities provided by language technologies. 2. To accumulate and enrich open, reliable, high-quality, reusable digital language resources and other digital language datasets. 3. To develop the language technology infrastructure, the application of language technologies in the public sector and public services, to create and improve publicly available information technology solutions and tools" (Gaidienė and Tamulionienė, 2022a).

The *European Language Equality* (ELE) project[1], funded by the European Commission, involved 24 EU official languages (including Lithuanian) and over 60 regional and minority languages. The project's primary goal was to develop the European Language Equality Programme[2], which includes a strategic research, innovation, and implementation agenda aimed at achieving full digital language equality in Europe by 2030. The programme was created collaboratively with the European Language Technology, Computational Linguistics, and language-centric AI communities, along with relevant initiatives, associations, and language communities. Research data were collected in 2021–2022, focusing initially on Lithuanian language resources and technologies, such as corpora, language models, computational grammars, and tools/services. The second stage identified Lithuanian stakeholders, projects, and funding related to language technology and AI. The conclusions of the ELE project were detailed in the Report on the Lithuanian Language (Gaidienė and Tamulionienė, 2022b).

The Lithuanian Artificial Intelligence Strategy: A Vision for the Future (2018) issued by the Ministry of the Economy and Innovation of the Republic of Lithuania. The aim of the strategy is "for Lithuania to become a regional leader on the basis of the existing resources, experience and potential. It aims to increase Lithuania's competitiveness among EU countries and to ensure its successful participation in the global AI ecosystem" (AI Strategy, 2018).

Building on the recent success of language technologies, largely driven by breakthroughs of technologies in generative artificial intelligence and large language models, European Commission established the *European Digital Infrastructure*

---

[1] More information about the project available at: *https://european-language-equality.eu/about/*

[2] See more: *https://european-language-equality.eu/about/*

Consortium Alliance for Language technologies (ALT-EDIC)[3] in 2024. The primary goal of this consortium is to facilitate the creation of large language models and the supporting ecosystem for all EU official languages. Lithuania is one of 17 founding members of the consortium. It is anticipated that this initiative will enhance European competitiveness in the field of generative AI. Moreover, the initiative is expected to have a significant impact on all areas of language research, including digital humanities.

The European Commission gave its green light to the *Lithuania's recovery and resilience plan "Next Generation Lithuania"* (Plan, 2021). Part of the plan is dedicated to the compilation and accumulation of Lithuanian language resources for development of Artificial Intelligence systems. The goals of this investment include the compilation of speech and text corpora, development of specialised language datasets, and digitisation of historical texts. All the compiled resources will be open and accessible for the public. It is hoped that the planned investment for 2024–2026 will give a great boost not only to the development of AI systems, but also to the development of digital humanities.

3. Lithuanian Heritage Platforms

Although there is a great number of institutions in Lithuania, including libraries and museums, that are dedicated to the storage and preservation of Lithuanian heritage cultural objects in different forms, we will present three most influential web heritage platforms that are solely dedicated to digitisation and preservation of cultural heritage in Lithuania, namely ePaveldas, the Database of Written heritage, and Lithuanian Folklore Archives database. These platforms are free and open to both the public and the science.

3.1. ePaveldas – Digital Cultural Heritage Platform

ePaveldas[4] is the largest digital cultural heritage platform in Lithuania. The main developer of the platform is Martynas Mažvydas National Library of Lithuania, its partners are as much as 24 institutions (libraries, museums, archives, institutes, etc.).

The content of the platform consists over 600,000 items of cultural heritage: 1) digitised 16th–20th century books in the Lithuanian language and Lithuanian publications; old Lithuanian and expatriate periodicals, cartographic publications; small print (decrees, proclamations, posters, programs, etc.); the court records of 1540–1845 of the Grand Duchy of Lithuania; church metrics on births, marriages and deaths; other documents reflecting different periods in the history of Lithuania; 2) exhibits of folk art and applied art; traditional folk sculptures; digitized audio recordings (piano rolls, shellac records) from museum collections, etc.[5]

The platform offers a range of services for users, e.g. to incorporate the items on the portal into users' exploration and learning processes, to conduct research or to use them in their creative work. These services include personalized virtual exhibition service,

---

[3] *https://language-data-space.ec.europa.eu/related-initiatives/alt-edic_en#the-alliance-for-language-technologies*
[4] *https://www.epaveldas.lt/*
[5] See more: *https://www.epaveldas.lt/about.*

sheet music player service, map service, virtual reality and augmented reality exposition service, etc.[6]

Important to note that the content of the platform is reflected in Europeana[7] and all items are labelled by Creative Commons licenses.
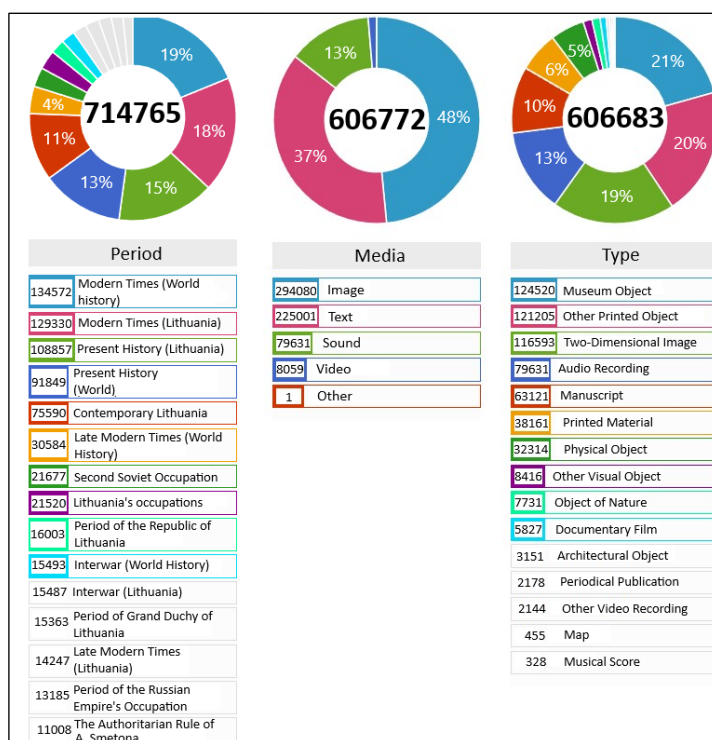

Figure 1. Structure of ePaveldas

3.2. Database of Written Heritage

The developer of the Database of Written Heritage[8] is the Institute of the Lithuanian Language. The corpus includes written texts from the first Lithuanian book (Martynas Mažvydas "Katechismuso prasti žodžiai", 1547) until the formation of standard Lithuanian at the end of 19th century. 155 sources currently available in the database[9].

Structure and content of the database: 1) the corpus (155 sources), 2) digital tools for text processing and language analysis (concordance generator, orthography changer, lemmatiser, etc.), 3) resources for language processing (internal lexicons of the morphology analyser, the structured database of lemmas and morphology, etc.).

---

[6] See more: *https://www.epaveldas.lt/about*

[7] *https://www.europeana.eu/lt*

[8] *https://seniejirastai.lki.lt*

[9] List of old writings: *https://seniejirastai.lki.lt/db.php*

Extended searches are available in the Database of Written Heritage, e.g. resource, author, year, place, etc. The Database of Written Heritage is periodically updated with new digitised resources.

In terms of value and significance, this database is the only one of its kind in Lithuania, as it provides invaluable material for cultural and historical digital research and serving as a widely utilised resource in academic studies (e.g. Ambrazas et al., 1998; Aleknavičienė, 2022).

3.3. Lithuanian Folklore Archives Database

The Lithuanian Folklore Archives Database[10] is the largest and the oldest Lithuanian folklore repository in Lithuania and globally. The developer of this database is the Institute of Lithuanian Literature and Folklore.

The database contains collections of the Lithuanian Science Society (working in 1907–1940 in Vilnius) and Lithuanian Folklore Archives (working in 1935–1939 in Kaunas), as well as materials from other sources.

The database includes around 3 million manuscript folklore texts, around 400,000 files of sound and video recordings, around 50,000 photographs and negatives, documents from the interwar Lithuanian Folklore Archives, personal manuscripts written by folklore collectors and researchers, etc.[11]. The Lithuanian Folklore Archives Database is periodically updated, filled with new resources. Extended searches are available in the Lithuanian Folklore Archives Database, e.g. title, type of document, collection, language, etc. The Lithuanian Folklore Archive provides rich material for cultural and literature research that are published in several cultural journals and books (e.g. Žarskienė and Nakienė, 2010).

3.4. eKultura.lt – Portal of Culture Heritage Digitisation

eKultura.lt[12] – informational-methodical portal of culture heritage digitisation. The developer of the portal – Lithuanian Museum Information, Digitization and LIMIS Center (LM ISC LIMIS), a branch of the Lithuanian National Art Museum. The portal intended for memory institutions and other legal and natural persons who carry out cultural heritage digitization activities, are preparing to do so or are simply interested in digital content created by others. Information can be found here: legal acts, standards, and recommendations relevant to digitizers of cultural heritage, links to virtual galleries created by Lithuanian memory institutions, etc.

3.5. Lithuanian Integral Museum Information System

Lithuanian integral museum information system (LIMIS)[13] – the virtual museum developed by the Lithuanian Art Museum. The system offers valuables and virtual products of Lithuanian museums – more than 100 institutions, 11 electronic services. The system is

---

[10] *https://www.tautosakos-rankrastynas.lt/lt/*
[11] See more: *https://www.tautosakos-rankrastynas.lt/en/*
[12] *http://www.ekultura.lt*
[13] *https://www.limis.lt*

searchable more than 800,000 cultural objects, browse constantly updated dictionaries of personalities and keywords; download or order high-resolution images, audio and video, 3D models; use digital content designated for public use for your own creative, educational, commercial or other activities. LIMIS offers a variety of search options.

4. Platforms of Lithuanian Language Resources

As the language and all the written and spoken sources are such an important part of digital humanities, not surprisingly there are numerous institutions, organisations, and private companies in Lithuania that accumulate, analyse, or share language resources in different forms and for different purposes. As there is no possibility to present the comprehensive list of all the available language resources, we will present the most prominent platforms and infrastructures in Lithuania that have the biggest impact on the development of digital humanities research.

4.1. CLARIN-LT

CLARIN (Common Language Resources and Technology Infrastructure)[14] is a distributed Pan-European research infrastructure designed to provide access to language data for researchers in fields such as linguistics, digital humanities, computer science, and related disciplines. The infrastructure is composed of interconnected centres located across various European research and data storage institutions.

Lithuania joined CLARIN ERIC in 2014 and subsequently established the CLARIN-LT consortium. This consortium currently comprises six institutions: Vytautas Magnus University, Vilnius University, Kaunas University of Technology, Klaipėda University, Mykolas Romeris University, and Baltic Institute of Advanced Technologies. CLARIN-LT[15] serves as one of the national centres within the broader CLARIN ERIC infrastructure and is coordinated by the Vytautas Magnus University in Kaunas.

The CLARIN-LT consortium operates the CLARIN-LT repository[16], which is dedicated to the storage and management of Lithuanian language resources. These resources primarily include corpora, lexical conceptual databases, and language processing tools (Petrauskaitė et al., 2022). The repository's mission is to adhere to FAIR principles (Findability, Accessibility, Interoperability, and Reusability) and promote open science. The CLARIN-LT repository serves multiple purposes: 1) storing data for research publications; 2) facilitating new language research; 3) supporting linguistic education, and 4) preserving language data, etc. Consequently, the data housed in this repository represents a valuable asset for researchers in digital humanities, providing essential resources for their academic investigations (e.g. see Dadurkevičius and Petrauskaitė, 2020; Rackevičienė et al., 2022).
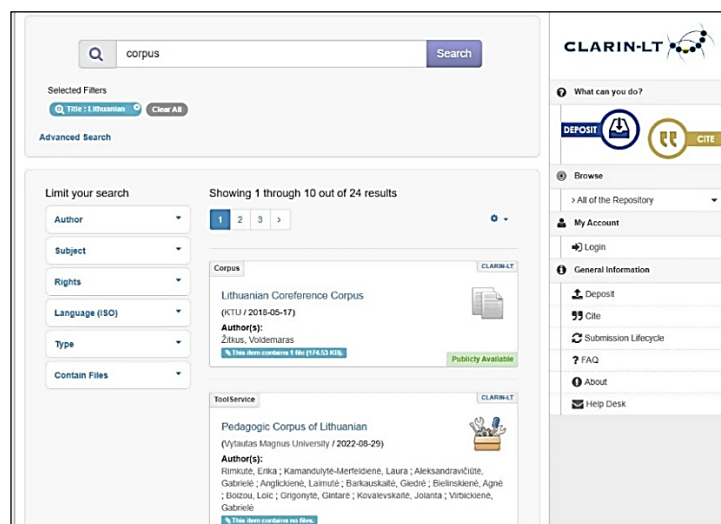
---

[14] *https://clarin.eu*
[15] *http://clarin-lt.lt/*
[16] *https://clarin.vdu.lt/xmlui/*

Figure 2. CLARIN-LT repository

4.2. The Republic of Lithuania Term Bank

The Republic of Lithuania Term Bank (LRTB)[17] is a state-managed information system that comprises a comprehensive collection of Lithuanian language terminology articles, organised according to established procedures and made accessible online.

The primary objective of the Term Bank is to ensure the consistent use of standardised Lithuanian terminology, particularly in the legislative documents of the Republic of Lithuania. It also aims to create a unified information system for various state institutions, which can be accessed and utilized by both legal entities and individuals. The resource is freely available to specialists across diverse fields, both within Lithuania and internationally. All terms in the Term Bank are categorised as approved (suitable for use in legislative acts), admitted (correct but not officially approved), and deprecated (incorrect). Currently, the Term Bank features more than 280 thousand terms.

The Term bank serves as a valuable resource for terminologists and language specialists.

4.3. Lithuanian Language Resources Information System "E. kalba" ("E-Language")

The Lithuanian Language Resources Information System "E. kalba" ("E-Language")[18] is a treasure trove of Lithuanian language resources. Developed by the Institute of the Lithuanian Language, "E. kalba" has been freely available to both professionals and the

---

[17] *https://terminai.vlkk.lt/*

[18] *https://ekalba.lt*

general public since 2021. The system provides access to 41 language resources, organised into three major groups: 1) e-dictionaries, 2) e-catalogues and databases, and 3) e-services. Among the resources available are the "Dictionary of the Standard Lithuanian Language", "Dictionary of the Modern Lithuanian Language", "Dictionary of the Lithuanian Language", "Database of Lithuanian Language Neologisms", "Dictionary of the Synonyms", "Lithuanian-Ukrainian Dictionary", "Catalogue of Riddles", "WordNet search", "Sentiment Analysis", "Word Formation Guide", and more.

One of the key advantages of "E. kalba" is the extensive data it offers. For example, the monolingual dictionaries include over 500,000 headwords and the e-catalogues contain 530,000 headwords, along with audio recordings of pronunciation and maps. Additionally, the system is regularly updated with new resources, ensuring its relevance and usefulness, e. g. two bilingual dictionaries are added to the system in 2024: "Little Ukrainian-Lithuanian Dictionary" and "Lithuanian-Ukrainian Dictionary".

"E. kalba" also features a multi-criteria search tool with an effective and sophisticated search mechanism, allowing for both quick and detailed searches across all resources or within specific ones (see more in Gaidienė, 2023). The system provides a user-friendly interface for viewing search results, which can be downloaded and saved in PDF and XML formats. It is also mobile-friendly, enhancing accessibility for us on the go.

This system is used not only by the public, but it also serves as a source for scientific purposes (see more Miliūnaitė, 2018; Murmulaitytė, 2022; Urbanavičienė, 2022).
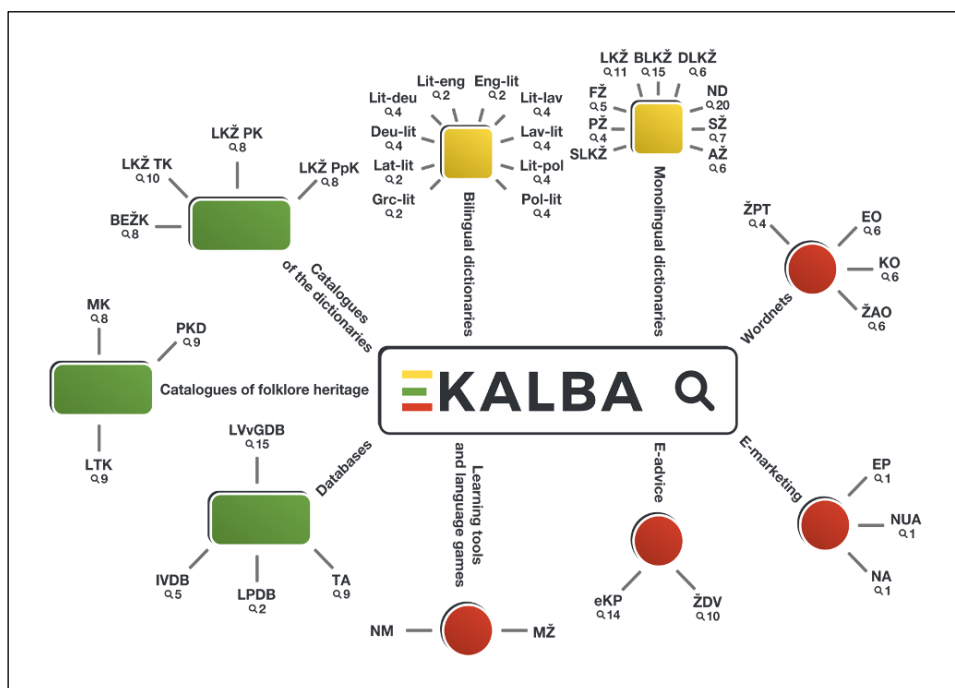

Figure 3. Structure of Lithuanian Language Resources Information System "E. kalba"

4.4. SITTI

The Institute of Digital Resources and Interdisciplinary Research (SITTI) maintains a platform[19], which encompasses a wide array of language resources, such as corpora, tools, datasets, dictionaries, frequency lists, word embeddings, and standards. While certain resources are also accessible through CLARIN-LT repository, there are specific resources and tools that are exclusively accessible via the SITTI platform. Additionally, the platform offers several widely-used language services tailored for the Lithuanian language researchers, such as concordance interface of the Corpus of the Contemporary Lithuanian Language and morphological taggers.

The newest service *Morfuoklis*[20]*,* an online tool for morphological analysis and synthesis of the Lithuanian language. *Morfuoklis* works on the basis of an open-source *Hunspell* platform solution (Németh, a), while its tokeniser and morphological analyser adapted from the `semantika.lt` portal. Currently, the *Morfuoklis* morphological knowledge base consists of 191,583 lemmas and 17,900 rules, divided into 4,943 groups. Most importantly *Morfuoklis* allows quick and easy morphological analysis of Lithuanian texts into three annotation standards, namely Jablonskis, MSD, UD/CoNLL-U, and Hunspell.

4.5. Spatial Data Infrastructure for the Lithuanian Language

The Spatial Data Infrastructure for the Lithuanian Language (SDI for LTL)[21] is being developed by the Institute of the Lithuanian Language. Initiated as a pilot project in 2019, the implementation of SDI for LTL is currently ongoing.

The primary objectives of SDI for LTL are to collect and process archival and contemporary geospatial linguistic data, provide tools and methodologies for analysing this data, support scientific insights into language phenomena, and offer mechanisms for publishing and disseminating geospatial linguistic information (see Čepaitienė, Mikulėnienė, 2022).

The structure of SDI for LTL will consist of three key elements: data, a portal, and adapted tools. The data will include: 1) internal data collected by scientists in the Institute of the Lithuanian Language ranging from scanned documents and simple tables to sophisticated databases; 2) external data provided by public institutions, such as the Centre of Registers or the State Data Agency; and 3) data provided by Esri itself and third parties.

In summary, as noted by Čepaitienė and Mikulėnienė (2022: 117), "The application of geographic information systems (GIS), which first emerged in the mid-20th century, has increased over the last few decades, enabling easier management of spatial and descriptive information – its collection, archiving, visualization, editing, integration, and analysis. Rapid development of GIS technologies (new infrastructure for storing and sharing spatial data, improved location identification, data collection, dissemination, and analysis), opened up new opportunities for their use only in geography but in many other scientific fields, dialectology being one of them."

---

[19] *https://sitti.vdu.lt/en/resources/*
[20] *https://sitti.vdu.lt/morfuoklis/lt*
[21] *https://maps.lki.lt/portal/home/*

4.6. Raštija.lt

Raštija.lt[22] is a portal for EU structural funds funded Lithuanian language projects implemented between 2012–2020: LIEPA (Speech recognition), SEMANTIKA (Corpora and NLP tools and services), VU MT (Machine translation system), LKIIS (some resources from the Institute of the Lithuanian Language). The developer of the portal is Vilnius University. The project has developed a web-based solution for the integration of existing, ongoing and future Lithuanian language and literary resources, products and services. The resources in the portal are described by general and specific ontologies and connected by a thematic network.

4.7. LKSSAIS (Semantika.lt) Tools

LKSSAIS[23] is a specialised language technology infrastructure and state information system that offers speech recognition and text analysis services for the Lithuanian language. The project was implemented by Vytautas Magnus University, which manages the information system, in collaboration with Kaunas University of Technology as a project partner.

The system features three language technology services and 12 associated IT solutions, all of which are freely accessible for internet users. The primary objective of LKSSAIS project was to compile Lithuanian text and speech corpora and apply natural language processing tools to deliver timesaving and efficiency-enhancing tools and electronic services.

4.8. Kalbu.vdu.lt, a Platform for Lithuanian Language Learners

Kalba.vdu.lt is an online platform of language resources for Lithuanian language learners developed at Vytautas Magnus University. The platform was developed within the framework of the project "Lithuanian Academic Scheme for International Cooperation in Baltic Studies". It features six digital resources, namely three corpora (pedagogic, learner, and morphologically annotated), the lexical database of Lithuanian language usage, pronunciation resources, and automatic accentuation tool *Kirčiuoklis*.

The aim of the provided resources is giving opportunities for learners of the Lithuanian language to observe the language by themselves and discover regularities, as well as irregularities of the language in authentic corpora. Learners are also provided with a modern automatic accentuation tool *Kirčiuoklis*[24], which assists users in correctly applying accents to provided texts. The accentuation tool is particularly popular among schoolchildren and learners of the Lithuanian language, as the Lithuanian accentuation system is notably complex and challenging to master.

---

[22] *https://raštija.lt*

[23] *https://semantika.lt/*

[24] *https://kalbu.vdu.lt/mokymosi-priemones/kirciuoklis/*

5. Digital Humanities Use Cases

The chapter provides an exploration of six digital humanities use cases and projects that address varied tasks within the field. These examples demonstrate the diverse applications of digital technologies in Lithuania's humanities research, ranging from diachronic semantic analysis to the preservation of cultural and linguistic heritage.

5.1. Diachronic Analysis

One of the challenges of digital humanities is detecting and representing semantic change of concepts in diachronic corpora. One of the technologies for representing the change of concepts over time could be linked data. The model called LLODIA (linguistic linked open data for diachronic analysis) (Armaselu et al. 2024) is based on existing vocabularies and approaches like OntoLex-Lemon, OntoLex-FrAC (McCrae et al., 2017; Chiarcos et al., 2022; Welty et al., 2006). LLODIA bridges concepts and resources within the diachronic analysis context in French, Latin, Hebrew, Lithuanian, and Romanian. The current resource allows the detection and representation of semantic change as linked data, and having in mind the ongoing progress in deep learning, digital humanities and undertakings in LL(O)D, ir promotes detecting and representing semantic change as LL(O)D which is an important topic for the future development of Semantic Web technologies (Armaselu et al. 2022).

5.2. Offensive Language

One of the digital humanities tasks is offensive language identification. One of the current projects for the Lithuanian offensive language identification (Oleškevičienė et al., 2023) started within the framework of COST action NexusLinguarum[25] examines the feasibility of identifying offensive language in Lithuanian based on the Simplified Offensive Language Taxonomy (SOLT) proposed by Barbara Lewandowska-Tomaszczyk (2022). The key feature of the proposed taxonomy is the complementation of existing offensive language ontologies and tagset systems aiming to integrate it into publicly accessible Linguistic Linked Open Data (LLOD) resources. The dataset used in the current study of offensive language identification in Lithuanian is a publicly available corpus of user-generated comments collected from two Lithuanian portals: *Delfi.lt* and *LRytas.lt* (Amilevičius et al., 2016).

5.3. Inside the Vilnius Ghetto

The digital humanities resource *Inside the Vilnius Ghetto: a musical excursion*[26] offers a virtual tour that explores the arts, education, and creative activities within the Vilnius Ghetto. This project provides an immersive 360-degree educational experience, allowing visitors to engage with the history of the former Vilnius Ghetto. It effectively combines dramatic history of a nation and a city with interactive elements of music and visualisation.

---

[25] *https://nexuslinguarum.eu/*
[26] *https://gete.lt/*

5.4. Miezinis Dictionary

Mykolas Miežinis dictionary "Lietuviszkai–Latviszkai–Lenkiszkai–Rusiszkai žodynas" published in 1894 includes about 15,000 words and is a valuable monument of Lithuanian lexicography. The original is digitalized at Indiana University[27]. The dictionary of M. Miežinis include Lithuanian words used in the places which the author lived such as author hometown (Vaitiekūnai, Radviliškis district), places where he studied at Grinkiškis, Dotnuva schools, Kėdainiai pro-gymnasium, Varniai priest seminary) and the places of his pastoral care (Šėta, Klykolaii, Didkiemis, Smalva, etc.). The resource is of key importance of carrying current comparative terminology studies (Kvašyte, 2022).

5.5. SLIEKKAS

Sliekkas[28] (Gelumbeckaitė et al., 2012) is a corpus of old Lithuanian writings, where the representation of the original spelling is transliterated into modern Lithuanian, followed by linguistic and morphological annotations. The corpus can be used for deep diachronic analysis. The lemmatized text is stored in a freely accessible annotated corpus (ca. 350,000 words), which includes 16th century religious literature and works by the Lithuanian national poet Kristijonas Donelaitis (1714–1780). The corpus provides an innovative scientific resource for historical and comparative linguistics. It is a rich dataset for exploring literary, religious and cultural settings in the Baltic countries, shedding light on the research related to the controversy between pre-Christian and Christian cultures (Kavaliūnaitė et. al. 2020).

5.6. A digital archive: Voices of spring

The Voices of Spring[29] is a digital archive created at Vilnius University. It includes a scientific edition of *The Voices of Spring* by Maironis and has a twofold aim of both an introduction to the classic poetry of Lithuanian literature, and a resource for linguistic research. The resource includes materials of the authorial process testimonials of creating, editing and publishing the poems. The digital archive includes all the autographs and editions created in Maironis's lifetime and the textual sources of musical compositions created after his poems.

6. Digital Humanities Future Trends

Although a considerable number of digital resources for the Lithuanian language and cultural heritage already exist, platforms such as *ePaveldas*, *E-language*, CLARIN-LT require continuous updates to address the evolving needs of Digital Humanities research and society. These updates should focus on expanding the variety of collections across

---

[27] *https://babel.hathitrust.org/cgi/pt?id=inu.30000132021498&seq=1*
[28] *https://titus.uni-frankfurt.de/sliekkas/*
[29] *http://www.pb.flf.vu.lt/*

different types of media and time periods, as well as developing and enhancing monolingual and bilingual corpora, dictionaries, lexicons, ontologies, and other linguistic resources.

The ongoing development of DH resources, tools, applications, and projects creates versatile research opportunities in fields such as comparative terminology, comparative linguistics, historical linguistics, and cultural studies. These research trajectories embrace religious, literary, cultural, and other societal domains at the same time providing grounds for preserving, enriching, and modernising cultural heritage and language.

Looking ahead, the digitization of all types of heritage objects is expected to intensify. Large-scale initiatives such as ALT-EDIC and *Next Generation Lithuania* are expected to facilitate the digitization and collection of vast amounts of contemporary Lithuanian language data for use in artificial intelligence (AI) systems. The increasing role of AI in the development and implementation of DH projects and services is likely to significantly enhance their scope and impact. These developments hold great promise for the DH field in Lithuania, offering more open and accessible data and tools, thereby supporting research, innovation, and the dissemination of cultural and linguistic heritage.

Acknowledgements

References

AI strategy (2018). *The Lithuanian Artificial Intelligence Strategy: A vision for the future*, available at *https://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_LT(1).pdf*.

Aleknavičienė, O. (2022). Lietuvių kalba kaip kalbos politikos objektas Prūsijos Karalystėje: XVIII a. pradžios diskursas (The Lithuanian Language as an Object of the Policy on Language in the Kingdom of Prussia: The Discourse of the Early 18th Century), Archivum Lithuanicum **24**, 57–102.

Ambrazas, S., Aleknavičienė, O., Zinkevičius, V. (1998). Istorinis lietuvių kalbos žodynas ir senųjų raštų kompiuterizavimas (Historical Dictionary of Lithuanian and Computerisation of Old Texts), Lietuvių kalbotyros klausimai **39**, 192–210.

Amilevičius, D., Petkevičius, M. (2016). *LITIS v.1*, CLARIN-LT digital library in the Republic of Lithuania, available at *http://hdl.handle.net/20.500.11821/11*.

Armaselu, F., Liebeskind, C., Marongiu, P., McGillivray, B., Oleškevičienė, G. V., Apostol, E. S., Truică, C. O., Gifu, D. (2024). LLODIA: A Linguistic Linked Open Data Model for Diachronic Analysis, *in* Chiarcos et al. (eds), *Proceedings of the 9th Workshop on Linked Data in Linguistics @ LREC-COLING* 2024, (25 May 2024, Torino, Italy). pp. 1–10.

Armaselu, F., Apostol, E. S., Khan, A. F., Liebeskind, C., McGillivray, B., Truică, C. O., Utka, A., Oleškevičienė, G. V., van Erp, M. (2022). LL (O) D and NLP perspectives on semantic change for Humanities research, Semantic Web **13**(6), 1051–1080.

Bennet, B., Fellbaum, C. (2006). *Formal Ontology in Information Systems*, Proceedings of the Fourth International Conference, FOIS 2006, Baltimore, Maryland, USA, November 9-11, 2006, IOS Press, Amsterdam.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., Jeffrey, S. (2016). *Digital Humanities*, MIT Press, Massachusetts.

Calzolari, N., Huang, C. -R., Kim, H., Pustejovsky, J., Wanner, L., Choi, K. -S., Ryu, P. -M., Chen, H. -H., Donatelli, L., Ji, H., Kurohashi, S., Paggio, P., Xue, N., Kim, S., Hahm, Y., He, Z., Lee, T. K., Santus, E., Bond, F., Na, S. -H. (eds) (2022). *Proceedings of the 29th International Conference on Computational Linguistics,* Conference (Oct 2022, Gyenongju, Republic of Korea), International Committee on Computational Linguistics.

Chiarcos, C, Apostol, E. S., Kabashi, B., Truică, C. O. (2022). Modelling frequency, attestation, and corpus-based information with OntoLex-FrAC, *in* Calzolari et al. (eds) *Proceedings of the 29th International Conference on Computational Linguistics*, (12-17 Oct. 2022, Gyenongju, Republic of Korea), International Committee on Computational Linguistics, pp. 4018–4027.

Chiarcos, C., Gkirtzou, K., Ionov, M., Khan, F., McCrae, J. P., Ponsoda, E. M., Chozas, P. M. (eds) (2024). *Proceedings of 9th Workshop on Linked Data in Linguistics @ LREC-COLING,* Workshop (25 May 2024, Torino, Italy).

Čepaitienė, A., Mikulėnienė, D. (2022). *Tarmynas – Lietuvos vietinių kalbos variantų duomenų bazės modelis. Ko ieškoti, ką galima rasti ir kaip lengva juo naudotis* (*Tarmynas – the Model of Lithuanian Local Language Variants Database Collecting, Searching and Applying Geolinguistic Data*). Lietuvių kalbos institutas, Vilnius. available at `https://lki.lt/wp-content/uploads/ 2022/09/Tarmynas-%E2%80%93-Lietuvos-vietiniu-kalbos-variantu-duomenu-bazes-modelis_galutinis.pdf.`

Dadurkevičius, V., Petrauskaitė, R. (2020). Corpus based methods for assessment of the traditional dictionaries, *in* Utka, A., et al. (eds) *Proceedings of the Ninth International Conference Baltic HLT 2020*, Kaunas, Lithuania, IOSS Press, Amsterdam, pp. 123–126.

Davidson, C. N. (2008). Humanities 2.0: Promise, perils, predictions, PMLA **123**(3), 707–717.

Doran, M. (2022). Introducing the Digital Humanities in Ireland Landscape Report Dataset. Studi irlandesi, A Journal of Irish Studies **12**(12), 33–42.

Fišer, D., Witt, A. (eds) (2022). *CLARIN: the Infrastructure for Language Resources*, De Gruyter, Berlin.

Gaidienė, A. (2023). Lietuvių kalbos išteklių informacinė sistema „E. kalba": nuo kūrimo istorijos, paieškos galimybių iki plėtros perspektyvų (Lithuanian Language Resources Information System "E. Kalba": from the History, Search Possibilities to Development Perspectives). Bulletin of Taras Shevchenko Kyiv National University. Literary studies. Linguistics, Folklore Studies **2**(34), 25–32.

Gaidienė, A., Tamulionienė, A. (2022a). European Language Equality in the Digital Age: the Case of Lithuania, Acta Linguistica Lithuanica **86**, 184–208, available at `https://journals.lki.lt/actalinguisticalithuanica/article/view /2164/2283.`

Gaidienė A., Tamulionienė A. (2022b). *Report on the Lithuanian Language*, available at `https://european-language-equality.eu/wp-content/uploads/2022/ 03/ELE___Deliverable_D1_23__Language_Report_Lithuanian_.pdf.`

Gelumbeckaitė, J., Šinkūnas, M., Zinkevičius, V. (2012). Senosios lietuvių kalbos tekstynas (SLIEKKAS) – nauja diachroninio tekstyno samprata (Old Lithuanian Reference Corpus (SLIEKKAS): A New Concept of a Historical Corpus), Darbai ir dienos **58**, 257–278.

Guidelines (2020). The Guidelines for the Development of the Lithuanian Language in the Digital Environment and the Progress of Language Technologies for 2021–2027, available at

`https://www.e-tar.lt/portal/lt/legalAct/`
`71152ab00eee11ebb74de75171d26d52`.

Hadalin, J. (2014). The Slovenian Digital Humanities Landscape–A Brief Overview, H-Soz-Kult 11.11.2014, available at
`https://www.hsozkult.de/debate/id/fddebate-132280`.

Helmi, L. C. (2021). Digital humanities: A paradigm for the 21st century, Culture and Human Behavior **2**(2), 5. DOI: 10.54729/2789-8296.1057.

Holm, P., Jarrick, A., Scott, D. (eds) (2015a). *Humanities World Report*, Palgrave Macmillan, London.

Holm, P., Jarrick, A., Scott, D. (2015b). The Digital Humanities *in* Holm et al. (eds), *Humanities World Report,* pp. 64–83.

Kavaliūnaitė, G., Kabašinskaitė, B., Kazakėnaitė, E. (2020). Renginys, skirtas senųjų tekstų skaitmeninimui (Event for Digitisation of Old Writings). Baltistica **54**(2), 361-365.

Kosem, I., Tiberius, C., Jakubíček, M., Kallas, J., Krek, S., Baisa, V. (eds) (2017). *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, Conference (19-21 Sept. 2017, Leiden, The Netherlands), Lexical Computing CZ s.r.o., Brno.

Kvašytė, R. (2022). Lietuvių ir latvių terminologija: istoriniai ryšiai ir gretinamieji tyrimai (Lithuanian and Latvian Terminology: Historic Connections and Comparative Studies), Terminologija **29**, 169-181.

Lewandowska-Tomaszczyk, B. (2022). A simplified taxonomy of offensive language (SOL) for computational applications, Konin Language Studies **10**, 213–227.

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P. (2017). The OntoLex-Lemon model: development and applications, *in* Kosem et. al. (eds), *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, Conference (19-21 Sept. 2017, Leiden, The Netherlands), Lexical Computing CZ s.r.o., Brno, pp. 587–597.

McPherson, T. (2009). Introduction: Media studies and the digital humanities, *Cinema Journal* **48**(2), 119–123.

Miliūnaitė R. (2018). Naujažodžių pateikimas ir paieškos galimybės Lietuvių kalbos naujažodžių duomenyne (Neologism Presentation and Search Possibilities in the Database of Lithuanian Neologisms), Bendrinė kalba **91**, 1–21.

Murmulaitytė D. (2022). Žodžių darybos vedlys plačiajai visuomenei ir specialistams (Word Formation Guide for the General Public and Specialists), *Lietuvių kalba,* **17**, 8–24, available at `https://www.zurnalai.vu.lt/lietuviu-kalba/article/view/` `31437/30396`.

Németh, L. (a). *hunspell*, availaible at `https://github.com/hunspell/hunspell`.

Nygren, T., Foka, A., Buckland, P. I. (2014). The status quo of digital humanities in Sweden: past, present and future of digital history. *H-Soz-Kult* 11.11.2014, available at `https://www.hsozkult.de/debate/id/fddebate-132280`.

Oleškevičienė, G. V., Selmistraitis, L., Utka, A., Gudelis, D. (2023). Offensive language in user-generated comments (*in* Lithuanian), Lodz Papers in Pragmatics **19**(2), 239–254.

Petrauskaitė, R., Amilevičius, D., Dadurkevičius, V., Krilavičius, T., Raškinis, G., Utka, A., Vaičenonienė, J. (2022). CLARIN-LT: home for Lithuanian language resources, *in* Fišer, D., Witt, A., (eds), *CLARIN: the Infrastructure for Language Resources*, De Gruyter, Berlin, pp. 511–536.

Plan (2021). Lithuania's recovery and resilience plan: Next Generation Lithuania, available at `https://commission.europa.eu/business-economy-euro/economic-` `recovery/recovery-and-resilience-facility/country-` `pages/lithuanias-recovery-and-resilience-plan_en#rrf-funded-` `projects-in-lithuania`.

Rackevičienė, S., Utka, A., Bielinskienė, A., Rokas, A. (2022). Distribution of Terms Across Genres in the Annotated Lithuanian Cybersecurity Corpus, *Respectus philologicus* **41** (46). Kaunas: Vilniaus universiteto Kauno fakultetas.

Svensson, P. (2010). The landscape of digital humanities, *Digital humanities quarterly* **4(1)**.

Toscano, M., Rabadán, A., Ros, S., González-Blanco, E. (2020). Digital humanities in Spain: Historical perspective and current scenario, Framing (Encuadrte) **29**(6).

Urbanavičienė J. (2022). Leksikografiniai tarmių duomenys lietuvių kalbos išteklių informacinėje sistemoje „E. Kalba" (Lexicographic Data on Dialects in the Lithuanian Language Resources Information System E-Language), *Lietuvių kalba*, **17**, 109–121, available at `https://www.zurnalai.vu.lt/lietuviu-kalba/article/`*`view/`* `31442/30408`.

Utka, A., Vaičenonienė, J., Kovalevskaitė, J., Kalinauskaitė, D. (eds) (2020). *Human LanguageTechnologies – The Baltic Perspectives* Proceedings of the Ninth International Conference Baltic HLT 2020, Kaunas, Lithuania, IOSS Press, Amsterdam.

Žarskienė, R., Nakienė, A. (2010). Lietuvių tautosakos rankraštynas: kaupimas, sisteminimas, skaitmeninimas (The Lithuanian Folklore Archives: Storage, Systematization and Digitalization), *Res Humanitariae* **7**, 183.

Welty, C., Fikes, R., Makarios, S. (2006). A Reusable Ontology for Fluents in OWL, *in* Benet, B., Fellbaum, C. (eds), *Proceedings of the Fourth International Conference*, FOIS 2006, Baltimore, Maryland, USA, pp. 226–236.