Glossaries

CyberResearch*

(including computer science, mathematical, and technological terms)

Abstract Vector

Any representation of a $\rightarrow$ sign that can be found in a sign list or a dictionary. These are called “abstract” because no sign in its native form would correspond to it 100%.

Doğu Kaan Eraslan, 297n54

$\rightarrow$ *features’ vector, Optical Character Recognition, sign*

Adab Corpus $\rightarrow$ see General Glossary

Adjacency Matrix

A matrix whose rows and columns are labeled with the $\rightarrow$ nodes, and whose cell value represents whether an $\rightarrow$ edge is present between the node labeled in the row and the node labeled in the column.

Doğu Kaan Eraslan, 306n87

It is essentially a table in which the rows and columns are cells from the corpus, and at the intersection of each row and cell, a number is given for how similar the two cells are.

M. Willis Monroe, 273

$\rightarrow$ *edge, node*

Administrative Texts $\rightarrow$ see General Glossary

AI $\rightarrow$ Artificial Intelligence

Akkadian $\rightarrow$ see General Glossary

* Although all words of an expression are capitalized in the list of terms below, when indicated after the “see also” arrow, words are capitalized when the expression has an acronym (for example, Conceptual Reference Model, CRM) but not capitalized when the expression does not have an acronym. There are some exceptions, such as Python, R, and Word2vec, that are applications or tools and thus always capitalized. The “definitions” provided in this glossary are neither exact nor comprehensive; rather, they are informative statements extracted from the chapters. In some cases, minor modifications (such as verb tense, or the addition of “it is”) have been made in order for the statement to make sense out of context. When multiple statements are presented for a term, they are listed according to the alphabetical order of their authors’ names. Cross-references point both to entries in this glossary as well as in the General Glossary.

Algorithm

A step-by-step procedure for solving a problem or accomplishing some end, especially by a computer.

Shannon Martino and Matthew Martino, 118n29

In Computer Science, an algorithm is a set of instructions a computer can execute to solve a (mathematical) problem.

Émilie Pagé-Perron, 197n16

→ *data mining*

American Standard Code for Information Interchange (ASCII) characters

Limited to characters that are used in the writing of English words without accents. They include but are not limited to: “3, b, R, a, m, k, [, ;”

Doğu Kaan Eraslan, 285n12

Analyse Logiciste

Analytical procedure developed in France focusing on outlining and highlighting the symbolic elements (and their relevant functions) that can describe the structure of an observed phenomenon. The logicist procedure is also a critical review of the methodologies and the epistemological perspectives adopted by different scholars.

Alessandro di Ludovico, 87n4

ANN → Artificial Neural Network

Annotation → see General Glossary

Application Programming Interface (API)

A set of rules and tools that defines how computers can interact. In this context, the → Online Cultural and Historical Research Environment API defines what database → items are available to extract and send to other programs.

Miller C. Prosser, 318n18

→ *item, Online Cultural and Historical Research Environment*

Aramaic → see General Glossary

Arbitrary Unit

A → spatial unit used as a primary spatial designator mostly on sites that lack any architectural remains or other clear spatial features that would allow for horizontal subdivision of an excavation area and defining vertical layers.

Sveta Matskevich and Ilan Sharon, 39

→ *spatial unit*

Artificial Adaptive Systems (AAS)

Biological computing methods, techniques, and → algorithms forming part of the vast world of → Natural Computation/Natural Computing, which is itself a subset of the → Artificial Sciences (AS).

Marco Ramazzotti, 61n5

→ *algorithm, Artificial Sciences, Natural Computing*

Artificial Intelligence (AI)

The development of models using symbol manipulation. The computation in the models is based on explicit representations that contain symbols organized in specific structures. The connectionist paradigm aims at massively parallel models that consist of a large number of simple and uniform processing elements interconnected with extensive links, the → Artificial Neural Networks .

Marco Ramazzotti, 62n9

→ *Artificial Neural Network*

Artificial Intelligence Model

Applied to → Mesopotamian Urban Revolution Landscape (MURL), Artificial Intelligence models are a type of artificial simulation of physical or cultural phenomena and features that have been previously encoded, so that from processing such models previously unknown relationships can be revealed through the features.

Marco Ramazzotti, 63n12

→ *Mesopotamian Urban Revolution Landscape*

Artificial Neural Network (ANN)

Artificial Neural Networks are a class of → algorithms for which the design is inspired by the observation of the mechanisms of the human brain. One of the most interesting features of ANNs is their ability to learn, which implies an attitude of flexibility, and to adapt to specific situations and datasets.

Alessandro di Ludovico, 94n26

They can be defined as Learning Systems; these are → algorithms for processing information that allow for the reconstruction, in a particularly effective way, of the approximate rules relating to a set of “explanatory” data concerning the considered problem (the input), with a set of data (the output) for which it is requested to make a correct forecast or reproduction in conditions of incomplete information.

Marco Ramazzotti, 64n16

→ *algorithm, Artificial Intelligence, neural networks, Self-Organizing Map, Word2vec*

Artificial Sciences (AS)

Sciences for which an understanding of natural and/or cultural processes is achieved by the recreation of those processes through automatic models.

Marco Ramazzotti, 61n5

→ *Artificial Adaptive Systems*

AS → Artificial Sciences

ASCII Characters → American Standard Code for Information Interchange characters

ASCII Transliteration Format (ATF)

ATF was created by the → Cuneiform Digital Library Initiative (CDLI) as a stable archiving format for the long-term storage of texts. It has evolved, first to adapt to the usage of CDLI, and later it branched out into → Oracc-ATF ... Whether one is working directly from the ancient → tablet or from paper publications, the ATF format is an excellent choice for encoding one's work for long-term preservation.

Émilie Pagé-Perron, 200

→ *American Standard Code for Information Interchange characters, Cuneiform Digital Library Initiative, Open Richly Annotated Cuneiform Corpus, tablet*

ASS → Artificial Adaptive System

ATF → ASCII Transliteration Format

Atomization

By atomization, we refer to the process of dividing data into many individual database → items. A text is atomized into many database items, each of which represents a single grapheme, either a letter or a → logosyllabic → sign.

Miller C. Prosser, 314n4

→ *item, logosyllabic, sign*

Attribute (for digital practice)

In the context of → Text Encoding Initiative (TEI), an attribute adds useful precision both for text analysis and for interpretation during the process of the exchange of data. An attribute stands within the element → tag.

Vanessa Bigot Juloux, 165

Attributes take the form of extra columns of information in the → nodes and → edges lists from which data can be filtered, such as by time period or transaction verb.

Emile Pagé-Perron, 220n85

→ *edge, node, tag, Text Encoding Initiative*

Attribute (in Archaeology) → see General Glossary

Auto-CM → Auto Contractive Map Neural Network

Auto-Contractive Map (Auto-CM)

An → Artificial Neural Network (ANN) that is characterized by a three-layer architecture consisting of 1) an input layer, where the signal is captured from the environment; 2) a hidden layer, where the signal is modulated inside the Auto-CM; and 3) an output layer, through which the Auto-CM feeds back upon the environment on the basis of the stimuli previously received and processed. The top layer is the input layer, and these are equivalent to the variables among which we seek to identify relationships.

Giulia Massini, in Marco Ramazzotti, 73n36

→ *Artificial Neural Network*

Bag-of-Words Model

A well-known method in textual analysis in which all of the individual lexemes from a text are flattened and placed in a large group. Syntax and order are lost and do not figure into the final analysis, but this model offers a level of simplicity that functions very well for comparing the similarity of texts within a corpus.

M. Willis Monroe, 270

→ *Continuous Bag-of-Words model*

Big Data

Large and complex datasets that require computational methods for their analysis.

Vanessa Bigot Juloux, 156n30

Bridge

→ Nodes that connect otherwise unconnected groups of nodes together ... Bridges consist of nodes that usually have a high → edge betweenness ... In the → Adab corpus, bridges are high-ranking individuals, or they represent homonyms (different individuals who have the same name and who should be represented with two or more separate nodes).

Émilie Pagé-Perron, 202, 212

→ *Adab corpus, edge betweenness, graph partitioning, node*

C(anonical)-ATF (C-ATF)

The backbone of the texts displayed by the → Cuneiform Digital Library Initiative (CDLI) ... It uses → American Standard Code for Information Interchange (ASCII) characters.

Doğu Kaan Eraslan, 285

→ *American Standard Code for Information Interchange characters, Cuneiform Digital Library Initiative*

C-ATF → **C(ANONICAL)-ATF**

CAL Code

Adopted by the Comprehensive Aramaic Lexicon (CAL) project, it is used for storing → Aramaic texts from the ninth century BCE to the thirteenth century CE. CAL Code was developed in the 1980s. It has changed considerably with the arrival of the internet and → Unicode, and it remains effective in giving a better understanding of displayed text. CAL Code uses → ASCII characters and was designed so that Aramaic scholars could use it without intensive training.

Doğu Kaan Eraslan, 288

→ *American Standard Code for Information Interchange characters, Aramaic, Unicode*

CBOW → Continuous Bag-of-Words

CDLI → Cuneiform Digital Library Initiative

CIDOC Conceptual Reference Model (CIDOC CRM)

A leading standard (→ ISO 21127/2006) for conceptual referencing of cultural heritage data. Developed in the last decade by one of the special-interest groups of ICOMOS for museums and object collections, it gained popularity in other domains of the humanities and cultural heritage. The CIDOC CRM was first adopted into the archaeological domain by English Heritage. After some adaptations and additions (the modified version was dubbed CRM-EH), they successfully mapped their field recording systems. Due to the standard's growing popularity, the initial group of researchers developed several extensions of the basic → CRM, aiming at incorporating additional information about the objects, such as bibliographic references, provenience, and other scientific analyses, as well as spatio-temporal models. The archaeological fieldwork extension, CRMarchaeo, accommodates stratigraphy-related concepts that were missing from CIDOC CRM.

Sveta Matskevich and Ilan Sharon, 45

The purpose of the CIDOC CRM is to establish a common framework for the sharing of data between GLAM institutions, and to function as an example of best practices in the cultural heritage domain.

A number of other → ontologies have already been incorporated into the CIDOC CRM ... with additional available extensions and use cases further enabling the representation of archaeological data and processes, the provenance of digital artifacts, and bibliographies. All of these provide opportunities for linking to other data streams. CIDOC CRM is a large and complex ontology consisting of 90 distinct entities and 149 property declarations.

Terhi Nurmikko-Fuller, 353–354

→ *Conceptual Reference Model, International Organization for Standardization, ontology*

CIDOC CRM → CIDOC Conceptual Reference Model**Clique**

An important concept in → network graph analysis, especially when working with → administrative records for which the source data generates undirected links among all of the people appearing in the same text. Each → tablet therefore forms a clique, that is, a fully connected subgraph where all → nodes are interrelated. Cliques are not exclusive: they can overlap and are formed by any group of nodes that have a link between each of them. For example, in a group where a, b, and c are all related, there are also three cliques of two nodes: (a, b), (a, c), and (b, c).

Émilie Pagé-Perron, 214

→ *administrative text, maximal clique, network graph analysis, node, tablet*

Cluster Analysis

A particular form of → multivariate analysis that divides a dataset into groups based on how near the data points are to each other and gives one the ability to analyze subgroups of the dataset using an → algorithm ... A type of analysis that divides a set of objects into groups (clusters) so that objects of one group are similar to each other, whereas objects within groups are dissimilar from the objects within other groups.

Shannon Martino and Matthew Martino, 118, 118n28

→ *algorithm, multivariate analysis*

Clustering Method

A statistical approach to group together → nodes that resemble each other. Those methods tell us about similarities in the dataset, but the researcher has to interpret the meaning of the groups ... Clustering methods have proven useful in → social network analysis for many decades.

Émilie Pagé-Perron, 220, 220n88

→ *cophenetic correlation coefficient, network analysis, node, social network analysis, Unweighted Pair Group Method with Arithmetic Mean*

Comma Separated Value (csv)

A format that most programs can interact with given the simplicity of its formatting and coding.

Shannon Martino and Matthew Martino, 137n62

A format that organizes data as a table, with new lines forming rows and commas delimiting columns.

Émilie Pagé-Perron, 206n49

Computer-aided Textual Analysis

A research method that can answer questions about themes or the usage of terms across a corpus containing millions of words ... Through computer-aided textual analysis, we can visualize what would otherwise be mere hunches gleaned from a close reading of the texts.

M. Willis Monroe, 258–259

Computer Science

Computer Science brings together disciplines including Mathematics, Engineering, the Natural Sciences, Psychology, and Linguistics.

Marco Ramazzotti, 60n2

Computer Semiotics

Any empirical approach mainly interested in the ways that humans and machines may communicate with each other using written languages, texts, and/or codes.

Marco Ramazzotti, 60n1

Computerized Data Modeling

The process of creating a digital → data model.

Sveta Matskevich and Ilan Sharon, 55n108

→ *data model*

Conceptual Reference Model (CRM)

Models that explain what things are in the digital world and how an object relates itself to the digital world.

Doğu Kaan Eraslan, 300n64

A *lingua franca* of recording systems, and the main purpose of ontological → mapping is to ensure compatibility of the terminologies used by various data management systems, which is one of the necessary conditions for creating → Linked Open Data clouds.

Sveta Matskevich and Ilan Sharon, 45

→ *CIDOC Conceptual Reference Model, data cloud, Database Management System, Linked Open Data, mapping, ontology*

Continuous Bag-of-Words (CBOW) Model

One of two alternative models one can choose from when training data in → Word2vec. While training, the CBOW model uses the various words that can appear in the same context as the target word for the prediction of semantic relationships.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 247n56

→ *bag-of-words model, Word2vec*

Continuous Skip-gram Model

One of two alternative models one can choose from when training data in → Word2vec. While training, the Continuous Skip-gram model predicts the words that may appear near the target word.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 247n57

→ *Word2vec*

Cophenetic Correlation Coefficient

A metric that can evaluate the validity of a particular → clustering technique together with a measure of distance or dissimilarity.

M. Willis Monroe, 274

→ *clustering method*

Correspondence Analysis

A → quantitative method used to describe a corpus of qualitative data. It falls within the statistical realm of “factor methods” and can be described as the analysis of the dependent relationships existing among the different features of the observed data. When the data are collected in a table, for example, the rows can be referred to as the observed specimens (in this case the seals’ scenes), while the columns refer to the different features (here, the different possible iconographic elements that appear in the scenes) that can pertain to the specimens. Correspondence analysis permits the investigation of the associations between rows and columns of the table and enables associations to be displayed in a graphic form.

Alessandro di Ludovico, 94–95n29

→ *hierarchical classification, quantitative analysis/method*

Cosine Similarity

A geometric method for determining similarity between rows in the → Document Term Matrix (DTM) by assigning each row to a → vector in multi-dimensional space and calculating the cosine of the angle between each vector.

M. Willis Monroe, 272

→ *Document Term Matrix, vector*

CountVectorizer

The function, CountVectorizer from the Scikit-learn module, runs through each bag and tallies a total list of all terms, then marks, for each bag, which terms appear in that bag. The end result is a matrix of all cells and all terms, with the count for each term in each cell tallied in the table.

M. Willis Monroe, 271–272

CRM → Conceptual Reference Model

Cuneiform → see General Glossary

Cuneiform Digital Library Initiative (CDLI)

Currently the major online database for → cuneiform texts. As the name implies, CDLI conserves epigraphical and semantic data of cuneiform texts, rather than for a specific language. It uses → American Standard Code for Information Interchange (ASCII) characters.

Doğu Kaan Eraslan, 285

The largest database of → cuneiform artifacts; it seeks to collect information about all → Mesopotamian inscribed objects. It stores this data as → metadata and images named “fatcrosses,” as well as → transliterations, normalizations, and translations in various languages ... One of CDLI’s roles is to maintain a complete catalog and digital copy of all cuneiform texts. Contributions are updated periodically by CDLI staff based on new research, changes in standards, and direct contributions from scholars.

Émilie Pagé-Perron, 199

→ *American Standard Code for Information Interchange characters, cuneiform, Mesopotamia, metadata, transliteration*

Cytoscape

A more powerful software than → Gephi. It is used not only by digital humanists but also by biologists and other scientists. It offers an extended number of features, especially for the statistical data analysis ... Cytoscape is free and open-source.

Émilie Pagé-Perron, 209, 218n80

→ *Gephi, graphing programs, weighted*

DANA → Digital Archaeology and National Archive

Data Cloud

A visualization diagram for a → triple store or multiple → graph datasets in the → Linked Open Data initiative.

Sveta Matskevich and Ilan Sharon, 49n95

→ *graph, Linked Open Data, triple store*

Data Mining

Such a task can be accomplished through a large number of different types of → algorithms and strategies. Typical of data mining is the search for correlations among features, like variables, within a certain dataset, so that the results of an investigation

of that dataset can give clues to predict the figures related to those features in similar cases, or the possible development of some features in case some other features change in a certain way. Just as resources can be found and extracted from a mine through proper types of investigation and extraction techniques, useful information that was not evident at a first look can be extracted from a dataset through data-mining techniques.

Alessandro di Ludovico, 91n21

→ *algorithm, Système Portable pour l'Analyse des Données, text mining*

Data Model

An abstract model that describes data structure and relations within a system, where the data is stored, and how it is processed.

Sveta Matskevich and Ilan Sharon, 48n91

A framework with rules. All databases have an underlying data model, which is simply an abstraction that defines how data is connected and processed in the database.

Miller C. Prosser, 319n20

→ *computerized data modeling*

Data Serialization

The process that enables us to save the data we are working on so that it can be reconstructed just as we had saved it in different platforms.

Doğu Kaan Eraslan, 299n61

→ *External Data Representation, JavaScript Object Notation*

Database Management System (DBMS)

A software that enables an administrator to create and manage databases and to monitor, modify, and analyze their interactions with users and other applications.

Sveta Matskevich and Ilan Sharon, 47n84

→ *Conceptual Reference Model, Identifier, meta-system*

DBMS → Database Management System

Deep Learning

A subfield of machine learning. It is characterized by the breakdown of complex representations into a hierarchy of simple components during the learning process.

Doğu Kaan Eraslan, 296n51

Dictionary

A type of abstract data structure (also known as an associative array or symbol table)

used in programming languages, implemented under the name `dictionary` in Python. A dictionary stores information as key value pairs. Use of dictionaries, at least in the context of Python, implies that we would have a non-negligible performance gain for operations in which we want to process a value by using its associated key.

Doğu Kaan Eraslan, 297n55

→ *map/mapping*

Digital Archaeology and National Archive (DANA)

Software for archaeological field recording developed by the Israel Antiquities Authority.

Sveta Matskevich and Ilan Sharon, 28n20

Distributional Semantic Models

In the context of → language technology, distributional semantic models keep track of the appearances of words according to their proximity to each other in order to measure their similarity. These models can be computed by counting the frequency of all the words in relation to their neighboring words or, more recently, by predicting the context where words appear.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 228–229

→ *language technology, Pointwise Mutual Information, Word2vec*

Distributional Semantics

The study of semantic similarities between linguistic → items by using mathematical methods to see how they are spread throughout a large dataset.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 229n17

→ *item*

Document-oriented Database

A data store, the main concept of which is the document. Documents can be organized hierarchically, grouped into collections by some criteria, or → tagged. Entities in a document store do not necessarily share a structure; differently structured documents can be stored in the same database.

Sveta Matskevich and Ilan Sharon, 51n99

→ *tag*

Document Term Matrix (DTM)

A large → “vector space” in which the cells reside. Each unique term within the corpus becomes a dimension on which the documents are plotted; for instance, if your corpus consists of documents that only use three words, the resulting vector space would be

recognizable as a three-dimensional plot. In practical terms the DTM is represented by a large table in which each row is one of the input documents, in this case a cell from one of our texts, and each column represents a unique term that shows up in the entire corpus.

M. Willis Monroe, 271

→ *cosine similarity, vector space*

DTM → Document Term Matrix

Edge

Connection between two vertices on a → graph.

Giulia Massini, in Marco Ramazzotti, 73n2

Connections between → nodes, often formed by shared terms, words, or other meaningful units of analysis.

M. Willis Monroe, 260

→ *graph, label, node*

Edge Betweenness

In → network analysis, a mathematical centrality measure of relationships between network data points.

Émilie Pagé-Perron, 212n62

→ *bridge, edge, network analysis*

Ego-network

A → network graph of one individual, and of its neighbors, and optionally all of the neighbors of these neighbors. In this context, a neighbor refers to a → node that is directly connected with the entity being investigated.

Émilie Pagé-Perron, 209n54

→ *Gephi, network graph, node*

Elamite → see General Glossary

Electronic Text Corpus of Sumerian Literature (ETCSL)

An online resource that provides access to the → transliterations (in both → Unicode and → American Standard Code for Information Interchange, ASCII) and translations of some 400 literary compositions, all dated to the late third or early second millennium BCE. The name encapsulates the form and function of this bi-part → web-based resource ... For the → Sumerian, there are transliterations rendered using the Latin alphabet and modern philological conventions (with no accompanying cuneiform, either

Unicode or otherwise) on a line-by-line basis. For the English translation, there are paragraphs of prose. The ETCSL provides external links to other related and complementary projects. These include other online corpora, resources for museological and palaeographical data, and a dictionary ... Although not actively developed since 2006, the ETCSL is a cited resource and in itself the focus of research. It was an innovative project in the realm of study of Sumerian literature in that it offered free and unrestricted online access to the data, with the aim of interlinking with external resources.

Terhi Nurmikko-Fuller, 351–352

→ *American Standard Code for Information Interchange characters, Sumerian, transliteration, Unicode, web*

Element

In the context of → Text Encoding Initiative (TEI), an element, which is a → markup → tag, is the first criterion to analyze data (in the form of text, image, sound, etc.). An element is conventionally marked up <element>. The markup data in between the opening tag "<" and the closing tag ">" indicates the type of information analyzed: for example, <persName>'Anatu'</persName>. One can easily understand that the element refers to a personal name, in this example 'Anatu ... Generally, the semantic of an <element> is very close in its abbreviation to its English vocabulary. For example, <interp> stands for "interpretation," <w> for "word," <l> for "line," and <text> for "text."

Vanessa Bigot Juloux, 165n68

→ *attribute, markup/markup tagging, tag, Text Encoding Initiative*

Encoding Scheme

A set of rules that standardizes the representation of an object in a digital environment.

Doğu Kaan Eraslan, 284n4

→ *interoperability/epigraphic interoperability, minimal semantic unit*

EpiDoc

Strictly speaking, EpiDoc is an → XML-scheme for adding → markup to ancient texts. Thus, it is applicable to all ancient languages, but it is predominantly used for Graeco-Roman corpora. It employs → Text Encoding Initiative (TEI) scheme as its backbone, and it accepts anything encoded with UTF-8. Since EpiDoc is fundamentally a markup language that tries to be applicable to a broad range of languages, it is fairly liberal when it comes to using values for attributes.

Doğu Kaan Eraslan, 289

→ *Extensible Markup Language, markup/markup tagging, Text Encoding Initiative, Unicode Transformation Format-8*

ETCSL → Electronic Text Corpus of Sumerian Literature

Event

A piece of contents (visual, textual, numeric measurement, for example) recorded by a particular observer at a particular time and place that conveys information concerning one or more semantic entities.

Sveta Matskevich and Ilan Sharon, 36

→ *syntactic elements*

Extensible Markup Language (XML)

Standardized in the late 1990s, Extensible Markup Language (XML) is a predefined → markup language that follows a standard syntax enabling data exchange to be easily → machine readable. One speaks then of → interoperability data—for example, XML would be a kind of artificial constructed language such as Esperanto, an artificial international language ... XML also offers to withstand time and new technologies deployed on the internet or locally (on a computer)—in other words, the current methodology is sustainable even as technology evolves.

Vanessa Bigot Juloux, 163–164, 164n63

A form of → markup language, like → HTML, which encodes → annotations, like the codes hidden behind the text we see when we use a word processing program, such as LibreOffice or Microsoft Word.

Emilie Pagé-Perron, 204n42

→ *annotation, data serialization, EpiDoc, HyperText Markup Language, interoperability/epigraphic interoperability, machine-actionable data/machine-readable data, markup/markup tagging, RDF-XML*

Extensible Markup Language (XML) Annotation

→ XML → annotations, including → Text Encoding Initiative (TEI), are perhaps the most popular means of annotating texts across disciplines. These types of annotations provide two major advantages: (1) ease of marking (with precision) the exact position of the annotated entity in the text and (2) flexibility of the → markup, making it possible to annotate multiple layers of information in one text. A good example of a successful annotated corpus available on the → web using XML is the → Electronic Text Corpus of Sumerian Literature (ETCSL).

Émilie Pagé-Perron, 204

→ *annotation, Electronic Text Corpus of Sumerian Literature, Extensible Markup Language, markup/markup tagging, Text Encoding Initiative, web*

External Data Representation (XDR)

Like → JavaScript Object Notation (JSON), it is a standardized serialization format.

Doğu Kaan Eraslan, 299n60

→ *data serialization, JavaScript Object Notation*

Factors of Variance

Also called factors of variation, they refer to elements that are believed to be constitutive of the observed state of that which is observed. This could include things such as elements of form (e.g., lines, angles), degrees of arc or convexity, color, and light intensity. Such features make the object distinguishable in a representation.

Doğu Kaan Eraslan, 308n93

→ *feature extraction algorithm*

Feature Extraction Algorithm

The process through which one extracts the → factors of variance from a digital representation of an object. For example, it could be an → algorithm that detects the corners and sides in a picture of a triangle.

Doğu Kaan Eraslan, 308n93

→ *algorithm, factors of variance*

Features' Vector

It is a convenient term for the mathematical construct that contains the patterns of the image of the → sign.

Doğu Kaan Eraslan, 297n54

→ *sign*

FRBR → Functional Requirements for Bibliographic Records

FRBROO

A formal → ontology for the representation of bibliographic information. It has been designed to merge with the → CIDOC Conceptual Reference Model (CIDOC CRM) and to facilitate the integration of museum and library data. It is based on the → Functional Requirements for Bibliographic Records (FRBR), which was originally designed as an entity-relationship model ... The first draft of FRBROO was completed in 2006, with the official publication of the version 1.0 at the end of 2009. It was approved and issued by January 2010 ... FRBROO is referred to in 60 of the CIDOC CRM's classes and in 55 properties.

Terhi Nurmikko-Fuller, 355

→ *CIDOC Conceptual Reference Model, Functional Requirements for Bibliographic Records, ontology*

Functional Requirements for Bibliographic Records (FRBR)

FRBR was developed independently of → CIDOC CRM, but coincidentally at approximately the same time (1991–1997). It was designed by a group appointed by the International Federation of Library Associations and Institutions (IFLA), and approved by IFLA Cataloguing Section in 1997. In 2003, an international working group began to examine the potential of merging FRBR with the CIDOC CRM. Unlike the CIDOC CRM, FRBR models products, not processes. It is smaller and less complex, with 52 entities and 64 properties. The combination of FRBR and the CIDOC CRM brought about → FRBROO—a new, object-oriented → mapping.

Terhi Nurmikko-Fuller, 355

→ *CIDOC Conceptual Reference Model, FRBROO, map/mapping*

Gaussian Distribution

A (bell curve) that has a central peak, is evenly distributed about the mean, and can be quantified with just the mean and the standard deviation (related to the width of the bell curve).

Shannon Martino and Matthew Martino, 121

Gephi

Gephi is free, open source, and easy to install ... Among the software available on the market, Gephi, with its intuitive interface, is a good entry-level tool for producing a first → network graph. Importing data into Gephi is straightforward, and it is easy to style → nodes and → edges depending on the user's needs. Gephi can also convert → graph data to numerous formats; this comes in handy when one is using more than one tool to manipulate the data. Gephi also integrates the sigma.js plugin, which can prepare a website on-the-fly for displaying an interactive network graph. Finally, Gephi includes a useful feature for isolating → ego-networks that provide information about individuals' relationships. An alternative to Gephi is → Cytoscape.

Émilie Pagé-Perron, 209, 209n55

→ *Cytoscape, edge, ego-network, graph data, graphing programs, network graph, node*

Glyph → see General Glossary

Graph

A mathematical abstraction that basically consists of a set of vertices and a set of → edges, where an edge represents the connection between two vertices on a graph.

Giulia Massini, in Marco Ramazzotti, 73

Used to visualize the connections between cells, a graph is made of → nodes and → edges, both of which are determined by the nature of the evidence.

M. Willis Monroe, 260

→ *edge, Gephi, graph data, node*

Graph Data

It enables the exploration of the relationships among entities.

Émilie Pagé-Perron, 197

→ *Gephi, graph*

Graph Model

A model that describes relations between objects as a collection of → nodes connected by → edges ... In a graph model, the data are structured so that each entity (subject), its attribute (object), and the relationships between them (predicate) create a → triple ... The main advantage of the graph model is its flexibility, which allows for querying large datasets by setting relations as query criteria.

Sveta Matskevich and Ilan Sharon, 46, 46n80

→ *edge, node, triple*

Graph Partitioning

Technique ... that consists of removing entities that act as a sole or faster point of access between groups in a → graph, thus separating these groups from each other. It may be evident from the visualization which individuals serve as → bridges, but using a → quantitative method to remove them is more efficient, since the threshold used will be numerical and not solely visual.

Émilie Pagé-Perron, 212

→ *bridge, graph, quantitative analysis/method*

Graph Visualization Software → Gephi

Graphical User Interface (GUI)

The medium in which users interact with the functions of the computer program through visual signs and symbols as opposed to written commands, as in the case of command line interface. Programs with GUI include, but are not limited to, Microsoft Word, Open Office Writer, Internet Explorer, and Mozilla Firefox.

Doğu Kaan Eraslan, 303n77

→ *Système Portable pour l'Analyse des Données*

Graphing Programs

Tools that allow one to visually interpret numerical data and export the results as images.

M. Willis Monroe, 274n36

→ *Cytoscape, Gephi*

GUI → Graphical User Interface

Hierarchical Classification

In the context of → correspondence analysis, it is based on similarity and proximity. It classifies groups and shows hierarchical dependences among them.

Alessandro di Ludovico, 95n29

→ *correspondence analysis*

Hierarchical Clustering Algorithm

Procedure of defining clusters (types) by considering how similar objects are to each other (in a quantitative fashion) and grouping “close” objects into the same type.

Shannon Martino and Matthew Martino, 120

→ *algorithm, cluster analysis, quantitative analysis/method*

Hierarchical Softmax

When using → Word2vec, a method for creating the output vector, the values for the words are stored in such a way that not all of the words have to be processed when calculating the probability of a word.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 247

→ *negative sampling, Word2vec*

HTML → HyperText Markup Language

HTTP → HyperText Transfer Protocol

HTTP URIs → HyperText Transfer Protocol Universal Resource Identifiers

Hub

A → node with an important number of connections compared to the average number of connections of nodes in a said → network graph.

Émilie Pagé-Perron, 210n58

→ *network graph, node*

Human Readable

A text that is destined to be read by humans, as opposed to → “machine readable,” which means it is intended to be processed by machines.

Doğu Kaan Eraslan, 287n17

→ *machine-actionable/machine-readable*

Hyperlinks

Often visualized as blue font and underlined text, hyperlinks allow users to move between → websites.

Terhi Nurmikko-Fuller, 338

→ *HyperText Transfer Protocol, web*

HyperText Markup Language (HTML)

HTML is used for formatting and the visual appearance of content of a → web page, but it does not capture meaning.

Terhi Nurmikko-Fuller, 339

→ *HyperText Transfer Protocol, web*

HyperText Transfer Protocol (HTTP)

The technology enabling interconnectedness over the → web through → hyperlinks. The acronym is familiar to us from the address bar displayed at the top of our browsers, starting with “http://”.

Terhi Nurmikko-Fuller, 338

→ *hyperlinks, HyperText Markup Language, HyperText Transfer Protocol Universal Resource Identifiers, web*

HyperText Transfer Protocol Universal Resource Identifiers (HTTP URIs)

These clusters of letters and characters form unique identifiers for each and every page on the → web. They are known as → Universal Resource Identifiers (URIs), and although other types of identifiers are possible, the ones we see when browsing are HTTP URIs. So common are the “http://” and “www.” prefixes that, when we discuss these resources, both are taken to be givens and thus are completely omitted; we refer to “Google,” for example, rather than to “http://www.google.com” ... For the most part, HTTP URIs point to specific pages. The majority of the content of these pages is text, encoded using the → HyperText Markup Language (HTML) ... On the → Semantic Web, rather than pointing to web pages, HTTP URIs serve as identifiers for specific data entities and information published on a page, a site, or in a resource.

Terhi Nurmikko-Fuller, 338–339

→ *HyperText Markup Language, HyperText Transfer Protocol, Semantic Web, Universal Resource Identifiers, web*

ID → Identifier

Identifier (ID)

In archaeological contexts, the identifier of an object is expressed in a label that consists of a combination of any alphanumeric characters attached to that object. Most data management systems (and most archaeological recording systems, be they manual or computerized) require that any entity has a unique ID.

Sveta Matskevich and Ilan Sharon, 36–37n52

A unique number that does not have a specific meaning. Its quality resides in its uniqueness.

Émilie Pagé-Perron, 204n45

International Organization for Standardization (ISO)

It is a non-governmental body that develops and publishes international standards of products, services, and systems.

Sveta Matskevich and Ilan Sharon, 45n72

It is the world's largest developer and publisher of international standards. It is an independent, voluntary organization, with representatives from each of the 162 member countries.

Terhi Nurmikko-Fuller, 353n55

Interoperability/Epigraphic Interoperability

The ability to transform the → encoding scheme of an encoded epigraphic phenomenon into another encoding scheme without losing data.

Doğu Kaan Eraslan, 284n5

→ *encoding scheme*

ISO → *International Organization for Standardization*

Item

An item could be an entire archaeological site or a single seed discovered in the course of excavation. An item can be anything from a place, to a person, to an image, to a text, to just about anything observable or conceptual.

Miller C. Prosser, 319

Iterative Process

The process of producing and analyzing data that generates results while moving toward an end research goal.

M. Willis Monroe, 270

Java Application Client

This client interface is a → Java Web Start application that launches on any computer with an internet connection.

Miller C. Prosser, 318

→ *Java Web Start, Online Cultural and Historical Research Environment, web*

Java Client → Java application client

Java Web Start

A technology that allows one to launch a computer program without going through an installation process.

Miller C. Prosser, 318n17

→ *Online Cultural and Historical Research Environment*

JavaScript Object Notation (JSON)

A serialization format, meaning that it allows us to save the state of our data. JSON is well standardized and well supported as an interchange format between different platforms.

Doğu Kaan Eraslan, 299n59

→ *data serialization, External Data Representation*

JSON → *JavaScript Object Notation*

k -plex

A relaxed → maximal clique in which k represents the number of connections that can be missing between → nodes when the nodes would still form a group, the k -plex. In a k -plex, most nodes are interconnected, but with a tolerance of k -number of absent → edges. In other words, an entity must be tied to all but k other entities in the group.

Émilie Pagé-Perron, 216

→ *edge, maximal clique, node*

Label

A form of → attribute used to identify → nodes and → edges easily instead of using only their unique → identifiers.

Émilie Pagé-Perron, 202n35

→ *attribute, edge, identifier, node*

Language Technology

A multidisciplinary field that studies and develops methods for processing human language with the help of computers.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 225n5

→ *distributional semantic models*

Lemma → see General Glossary

Lexicography → see General Glossary

Library

A collection of commonly used programming functions that are distributed as a package for use by a wide range of users.

M. Willis Monroe, 268n29

In cyber-research contexts, “library” has two distinct usages: 1) a “digital library” is an online collection of digital objects that can be surrogates of actual objects or parts of objects, or born digital objects (i.e., → Cuneiform Digital Library Initiative) 2) a “software library” is a bundle of code that has a specific focus of application and that can be reused by many while developing software (i.e., Python library).

Emilie Pagé-Perron, 218n79

→ *Cuneiform Digital Library Initiative*

Linked

In the world of the digital humanities, the term “linked” can have a specific connotation, meaning the mode of modelling data for sharing across the → web with other datasets with which one’s data was previously not connected. Within the → Online Cultural and Historical Research Environment (OCHRE) system, the term refers to database → items that “point” to each other, thereby creating a link between them. The database is a network of millions of files pointing at each other, i.e., “linked” to each other.

Miller C. Prosser, 319n23

→ *item, Online Cultural and Historical Research Environment, web*

Linked Data (LD)

Information that has been captured and published using the → w3c standards of → Resource Description Framework (RDF) and → HyperText Transfer Protocol (HTTP) URIs. The term does not, however, signify that the data in question has been made openly available.

Terhi Nurmikko-Fuller, 344

A directed → network graph in which all → edges are named and directional.

Émilie Pagé-Perron, 201n30

→ *edge, HyperText Transfer Protocol Universal Resource Identifiers, linked, network graph, ontology, Resource Description Framework, Simple Knowledge Organization System, World Wide Web Consortium*

Linked Open ‘Data (LOD)

It combines both → Linked Data (LD) and → Open Data (OD): it is often based on OD that has been converted to → Resource Description Framework (RDF), and it is similarly available for consumption and reuse without restrictions or limitations.

Terhi Nurmikko-Fuller, 344

→ *linked, Linked Data, Open Data, Resource Description Framework*

Locus → see General Glossary

LOD → Linked Open Data

Logarithm

By common definition, a logarithm (abbreviated “log”) is a quantity representing the power to which a fixed number (the base) must be raised to produce a given number.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 239n44

→ *Pointwise Mutual Information*

Logosyllabic → see General Glossary

Machine-actionable/Machine-readable

A text intended to be processed by machines.

Doğu Kaan Eraslan, 287n17

Machine-actionable Data/Machine-readable Data

Data structured in a way that it can be processed by computer software.

Émilie Pagé-Perron, 200n27

→ *scraper*

Manuel de Codage (mdc)

The standard developed for processing ancient Egyptian hieroglyphs. mdc, specifically its last version, mdc-88, has seen a lot of variants, and it uses → American Standard Code for Information Interchange (ASCII) characters.

Doğu Kaan Eraslan, 287

→ *American Standard Code for Information Interchange characters*

Map/Mapping

In the context of a → dictionary-data structure, it is to assign a value to a key.

Doğu Kaan Eraslan, 297n55

→ *dictionary*

Markup/Markup Tagging

Marking up online content can be compared to annotating a hard-copy text by hand (i.e., writing notes with a pen on the paper manuscript)

Vanessa Bigot Juloux, 163

→ *annotation, element, Extensible Markup Language annotation, HyperText Markup Language, tag*

A way to annotate information using → tags. → HyperText Markup Language (HTML) is a markup language.

Émilie Pagé-Perron, 204n43

Maximal Cliques

They are the largest → cliques identifiable in a network. If a text mentions “a” and “b,” and another text mentions “a,” “b,” and “c,” then “a, b, c” forms a maximal clique, overlapping with the “a-b” clique.

Émilie Pagé-Perron, 214

→ *clique, k-plex*

Maximally Regular Graph (MRG)

In the context of the → Mesopotamian Urban Revolution Landscape (MURL) study, MRG represents an alternative graph-theory representation of the underlying power network with respect to the one obtained by simply reconstructing the geography of common board affiliation ... From a → Minimum Spanning Tree (MST), generated from any metric, the MRG reshapes the links among nodes in order to maximize the fundamental and the most regular structures implicated in any dataset ... Compared to the MST, therefore, the MRG adds all (and only) those extra features that are really useful in understanding the prototypes that are hidden in the database; in other words, it adds the optimal amount of complexity that is necessary to read the phenomenon.

Marco Ramazzotti, 67n28

→ *Mesopotamian Urban Revolution Landscape, Minimum Spanning Tree*

Maximally Regular Graph (MRG) Algorithm

A new type of semantic → graph that uses a new index for detecting structural/topological complexity information in undirected graphs (H function) ... The MRG → algorithm generates, starting from the → Minimum Spanning Tree (MST), the graph presenting the highest number of regular microstructures that make use of the dataset's most important connections.

Marco Ramazzotti, 67n28

→ *algorithm, graph, Maximally Regular Graph, Minimum Spanning Tree*

Mesopotamian Urban Revolution Landscape → see General Glossary

Metadata

Information about the text that is external to the textual information itself and can be used for classification purposes. It may describe provenience, period, genre, etc.

Émilie Pagé-Perron, 199n19

Metadata provides additional information about the text in question. In Assyriology, this information includes date and provenance.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 225n8

→ *Cuneiform Digital Library Initiative, Open Richly Annotated Cuneiform Corpus*

Meta-model

A → model that describes a model. More specifically, it is a high-level abstraction that uses modeling language to describe a model.

Sveta Matskevich and Ilan Sharon, 35n51

→ *model*

Meta-system

Data description system external to a data management system; a system that describes a system.

Sveta Matskevich and Ilan Sharon, 38n55

→ *Database Management System*

Metric Scaling

It is the process of considering each → item as a point in a multidimensional space and calculating the “distance” between the points using a metric, just as you find the distance between points on a map.

Shannon Martino and Matthew Martino, 120

→ *item*

Minimal Semantic Unit

The smallest distinctly perceived element. This corresponds to the smallest delimited element in the case of an → encoding scheme, since the perception is done by the machine.

Doğu Kaan Eraslan, 286n14

→ *encoding scheme*

Minimum Spanning Tree (MST)

From a conceptual point of view, the MST represents the energy-minimization state of a structure. In fact, if we consider the atomic elements of a structure as the vertices of a → graph and the strength among them as the weight of each → edge linking a pair of vertices, the MST represents the minimum of energy needed so that all the elements of the structure preserve their mutual coherence. In a closed system, all the components tend to minimize the overall energy. So the MST, in specific situations, can represent the most probable state for the system.

Marco Ramazzotti, 67n28

→ *edge, graph, Maximally Regular Graph*

Minimum Spanning Tree (MST) Data Mining

A method that permits branching structures that also reveal clustering in archaeological datasets; this method has been applied as a model for → spatial analysis in different ways.

Marco Ramazzotti, 67n28

→ *data mining, Minimum Spanning Tree, spatial analysis*

Mixed Method

→ Quantitative and → qualitative methods combined.

Vanessa Bigot Juloux, 163

→ *qualitative method, quantitative analysis/method*

Model

It is as an external and explicit representation of a system or a part of reality.

Sveta Matskevich and Ilan Sharon, 35n51

→ *meta-model*

MRG Algorithm → Maximally Regular Graph algorithm

MST → Minimum Spanning Tree

MST Data Mining → Minimum Spanning Tree data mining

Multivariate Analysis

One method often utilized by archaeologists to create typologies, but it is not simply one tool that is easily employed in the same way by everyone regardless of the purpose. It is, rather, a general description for a type of analysis that takes many forms. One must always adapt the method to the material that one wishes to analyze. Using a multivariate analysis, archaeologists can determine the types that define a typology through the analysis of multiple variables and their interdependency.

Shannon Martino and Matthew Martino, 118

MySQL

A → relational database software that employs the → Structured Query Language (SQL) to fetch data.

Émilie Pagé-Perron, 196n7

→ *querying, relational database, Structured Query Language*

Namespace

In → EpiDoc, it refers to the appellation of the → tags with relation to the information they convey. There could be several options in determining which tag should be used for a personal name, such as John Doe Smith. We could, for example, decide to use “name,” “personalName,” or both. Which appellation better frames the interpretation of the name depends on the context, and, thus, on the project, although it should be kept in mind that the more interpretation one provides, the more time one must invest in constructing a coherent namespace.

Doğu Kaan Eraslan, 302n72

→ *EpiDoc, tag*

Natural Computing (NC)

Nonlinear dynamics and pattern formation computing inspired by concepts, principles, and mechanisms underlying natural systems.

Marco Ramazzotti, 60n2

→ *Artificial Adaptive System, Natural Computing algorithms*

Natural Computing (NC) Algorithms

Inspired by natural and biological phenomena, they include evolutionary → algorithms, → neural networks, molecular computing, and quantum computing.

Marco Ramazzotti, 60n2

→ *algorithm, Natural Computing, neural networks*

Natural Language Processing (NLP)

A broad term that refers to methods for teaching computers to understand human language. At the risk of over-simplifying, the idea is that after a computer has processed a large number of words in a text, it should be able to extract meaning from a text it has never seen. Most work in this field has focused on teaching computers to understand English. The benefits of NLP have not yet reached the study of the languages of the ancient world.

Miller C. Prosser, 322n29

NC → Natural Computing

NC Algorithms → Natural Computing algorithms

Negative Sampling

When using → Word2vec, a method for creating the output vector, which enables a procedure in which only a small part of the model's weights is updated.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 247

→ *hierarchical softmax, Word2vec*

Network Analysis

It permits one to approach the data in large amounts of texts through a new lens, by inquiring quantitatively with a focus on the relationships among entities of interest. Such inquiries facilitate the detection of meaningful patterns that could not be easily perceived using traditional methods ... The use of network analysis is well established in some fields, such as biology and sociology, but it is only slowly gaining popularity among Assyriologists ... In textual studies, network analysis is a digital methodology that focuses on the relationships among entities in the written record and enables these relationships to be studied at a larger scale than is normally feasible with traditional philological methods.

Émilie Pagé-Perron, 194, 194n3, 196

→ *quantitative analysis/method*

Network Analysis Algorithms

→ Algorithms specifically geared to → network analysis.

Émilie Pagé-Perron, 197n16

→ *algorithm, network analysis*

Network Graph

The total of all data points, the → nodes, and their relationships—i.e., the → edges. Some nodes can be completely disconnected from others and form independent ensembles. Together, connected and disconnected nodes form the network graph ... Network graphs are composed of nodes and links; nodes are data points representing entities. Edges link nodes, representing the relationships between those nodes. This data takes the form of → “triples” comprising two nodes and an edge that connects them.

Émilie Pagé-Perron, 201, 201n29

→ *edge, graph, node, triple*

Network Graph Analysis

→ An administrative text ... is an ideal candidate for → network graph analysis in which individuals who appear in the text are chosen as entities that will become nodes in the final → graph, and their co-occurrence generates links that connect them together.

Émilie Pagé-Perron, 201

→ *administrative text, graph, network graph*

Network Graph Triples

In a → network graph, data takes the form of → “triples” comprising two → nodes and an → edge that connects them.

Émilie Pagé-Perron, 201

→ *edge, network graph, node, triple*

Network Graph Visualization

When using a corpus in which the texts, such as sub-archives discussing only one type of activity, are homogenous, network graph visualization is an efficient means for giving an overview of the relationships without requiring extensive manipulation of the → graph.

Émilie Pagé-Perron, 210

→ *graph, network graph*

Network Theory

The study of systems and patterns found in → network graphs. → Social network analysis uses network theory to understand social relationships.

Émilie Pagé-Perron, 195n5

→ *network graph, social network analysis*

Neural Networks

A computational model inspired by the human brain where interconnected nodes (neurons) work in parallel to find out how to solve a problem by themselves when given an example.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 229n19

→ *Artificial Neural Network, Natural Computing algorithms*

NLP → Natural Language Processing

Node

In many cases, nodes represent texts within a corpus or documents within a genre.

M. Willis Monroe, 260

In → network analysis, nodes represent entities that can be of one or more types, such as people, institutions, or places.

Émilie Pagé-Perron, 201

In the context of → Mesopotamian Urban Revolution Landscape (MURL), nodes are the spatial relations between points (sites) ... Each distance between a pair of nodes becomes the → weighted → edge between this pair of nodes.

Marco Ramazzotti, 66, 73

→ *edge, edge betweenness, ego-network, Mesopotamian Urban Revolution Landscape, network analysis, Self-Organizing Map, weighted*

Normalization

The process of expressing different phenomena with a common unit. For example, you cannot add apples to oranges, but you can add fruits to fruits. Thus, in a context in which we are interested only in adding fruits, since apples and oranges are fruits, we can add apples to oranges. The passage from apples to fruits is a case of the application of a normalization procedure.

Doğu Kaan Eraslan, 293n39

Normalized Pointwise Mutual Information (NPMI)

An extended version of the $\rightarrow$ Pointwise Mutual Information (PMI) formula, which removes some of the low-frequency bias and gives the score fixed upper and lower boundaries.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 240
 $\rightarrow$ *logarithm, Pointwise Mutual Information*

NoSQL

$\rightarrow$ Database models that utilize a non-tabular or not only tabular data structure. Examples include a document store (such as $\rightarrow$ Extensible Markup Language), $\rightarrow$ triple store ($\rightarrow$ Resource Description Framework), graph databases, and object-oriented databases.

Sveta Matskevich and Ilan Sharon, 51n98
 $\rightarrow$ *database model, Extensible Markup Language, Resource Description Framework, triple, triple store*

NPMI $\rightarrow$ Normalized Pointwise Mutual Information

OCHRE $\rightarrow$ Online Cultural and Historical Research Environment

OCR $\rightarrow$ Optical Character Recognition

OD $\rightarrow$ Open Data

OM $\rightarrow$ OntoMmedia

Online Cultural and Historical Research Environment (OCHRE)

The OCHRE database runs on a server professionally supported by the Digital Library Development Center at the Regenstein Library on the University of Chicago campus. All core data is stored in this database. Users access data through the OCHRE $\rightarrow$ Java application client. This client interface is a $\rightarrow$ Java Web Start application that launches on any computer with an internet connection. Because OCHRE is an online-database

environment, project members from anywhere in the world have access to data live and in real time ... One of the simple principles underlying the OCHRE → data model is that each discrete meaningful unit of observation is a separate database → item. Each database item is stored as an → Extensible Markup Language (XML) file ... Millions of individual XML files are related one to another through a variety of organizational methods, with the primary among these being the hierarchy.

Miller C. Prosser, 318–319

→ *data model, Extensible Markup Language, item, Java application client, Java Web Start, linked*

Ontology

Formalized structures that form part of expert systems. They constitute the underlying data infrastructure for projects, tools, and datasets published as → Linked Data ... Ontologies are used to define the types of entities that occur in the dataset (classes) and the possible relationships between those entities (properties) define the type of entity a subject or object is, and the properties describe predicates, including the directionality of the relationship it captures. The class from which the property runs is the domain; the class to which it runs is the range.

Terhi Nurmikko-Fuller, 347–348

→ *Linked Data, upper ontology*

OntoMedia (OM)

Like the → CIDOC Conceptual Reference Model (CIDOC CRM), OM is event-based. It focuses on the representation of the narrative in multi-media and has been designed as linkable to the CIDOC CRM. The aim of this → ontology is to enable the human-like, vague questions we might use in conversation when trying to identify a given story ... OM represents the narrative content of heterogeneous media. It is based largely on two interlinked topics: the literary genres of fantasy and science fiction and the fan-fiction associated with them, both manifesting as a number of genre-specific classes.

Terhi Nurmikko-Fuller, 357

→ *CIDOC Conceptual Reference Model, ontology*

Open Data (OD)

Data that has been made publicly available without access restrictions resulting from copyrights or paywalls. It often manifests as raw, or largely unanalyzed datasets, made downloadable in formats such as → Comma Separated Value (csv) files.

Terhi Nurmikko-Fuller, 344

→ *Comma Separated Value, Linked Data, Linked Open Data*

Open Richly Annotated Cuneiform Corpus (Oracc)

Oracc presents itself as a platform hosting sub-projects that are managed by their contributing teams. Oracc uses standardized encoding schemes that are based on the same original format: → ASCII Transliteration Format (ATF).

Émilie Pagé-Perron, 200

One of the largest corpora of → Sumerian and → Akkadian texts, consisting of over seventeen thousand texts (almost two million words). Roughly half of it has been annotated ... In general, in the Oracc corpus, the → metadata added to the texts has been done during different projects over a number of years. This is why some texts have more metadata than others. Some texts include information on provenance, the period when they were written, and genre. Individual words may have such → tags as transcription, dictionary form, translation, part of speech, and language.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 225, 228
→ Akkadian, annotation, ASCII Transliteration Format, metadata, Sumerian, tag

Optical Character Recognition (OCR)

A subfield of computer vision that permits machines to extract text characters from a pixel-based image ... OCR has become a viable option to use in epigraphy largely due to the popularization of → deep-learning → algorithms. The passage, however, between the → raster image and → vector graphic, which can be stored in encodings, is actually quite tricky ... The OCR process would actually generate a → features' vector that could be compared against the → abstract vector of the → sign.

Doğu Kaan Eraslan, 296, 296n50, 297

A process by which a machine scans a typed or handwritten document and converts the text into digital characters.

Miller C. Prosser, 323n30

→ abstract vector, algorithm, deep learning, features' vector, raster image, sign, vector, vector graphic

Oracc → Open Richly Annotated Cuneiform Corpus

Oracc-ATF

Specific characteristics of Oracc-ATF include a wider array of characters permitted in the → transliterations lines and the validation of data content managed at the project level. As a result of these characteristics, transcription standards vary.

Émilie Pagé-Perron, 200

→ ASCII Transliteration Format, Open Richly Annotated Cuneiform Corpus, transliteration

Physical Medium → see General Glossary

PMI → Pointwise Mutual Information

Pointwise Mutual Information (PMI)

A statistical method used to find collocations and associations between words. It measures the reduction of uncertainty about the occurrence of *word₁* when *word₂* is known to be present within a given distance. Thus, we may be able to generate groups of words that are syntagmatically related to each other ... PMI is a count model whereas → Word2vec uses so-called → Artificial Neural Networks (ANNs) to predict word co-occurrence.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 229, 238
→ *Artificial Neural Network, Word2vec*

Pre-processing

The task of preparing the (textual) data before starting the → annotation process or some computational analysis.

Émilie Pagé-Perron, 197m15
→ *annotation*

Protocol

A set of rules for the exchange or transmission of data between machines.

Terhi Nurmikko-Fuller, 340n8

Qualitative Method

Qualitative methods enable the collection of data by observation through the researcher's participation in the investigation; in order to reflect the purpose of the analysis, the most relevant data is used for a deeper investigation, such as parsing with → R. This data is recorded and sorted according to its nature in order to match categorical data. Data can be classified into two types: numerical and categorical data. Numerical data is used → quantitatively to measure the numerical values' outcome, and categorical data is used qualitatively to organize nominal values into categories. Categorical data is used for → text and → data mining processes. Qualitative methods do not exclude the quantitative methods because, in some cases, calculations based on significant data types are needed.

Vanessa Bigot Juloux, 162–163
→ *data mining, mixed method, quantitative analysis/method, R, text mining*

Quantitative Analysis/Method

Quantitative methods allow the researcher to count and measure a group of data; they use mathematical formulas and statistics to express results in terms of numbers or sets of numbers. Graphics or templates are often used to communicate results.

Vanessa Bigot Juloux, 162

Although Assyriologists are generally not acquainted with quantitative analysis methods, these techniques are making a remarkable entry into the field ... using quantitative analyses is helpful for Assyriological studies: preparing data in a → machine-actionable way enables researchers to explore a corpus using a large array of alternative and complementary approaches, from simple statistical models to machine learning → algorithms. The results of such methods, including → network analysis, yield exact results that can be counted, compared, and used as strong evidence to build an argument.

Émilie Pagé-Perron, 216–217

→ *algorithm, correspondence analysis, graph partitioning, machine-actionable/machine-readable, mixed method, network analysis, qualitative method*

Quantitative Data

Countable and measurable specimens on which mathematical operations, such as statistical analysis, can be performed.

Émilie Pagé-Perron, 218n78

Querying

The processes of extracting information from a database based on the user's input.

Doğu Kaan Eraslan, 285n9

→ *MySQL, Structured Query Language*

R

An open-source software environment, mostly used for statistics, graphics, and data manipulation. One of the major benefits of R is its popularity, which guarantees regular developments by a broader community of R users.

Vanessa Bigot Juloux, 182n127

→ *qualitative method*

Ras Šamra → see General Glossary

Ras Shamra Tablet Inventory (RSTI)

RSTI is an → OCHRE research project co-directed by Miller C. Prosser and Dennis Pardee. A primary goal of the project is to create reliable digital editions of the texts in the

Ras Šamra-Ugarit corpus within a → research database environment. More than just a data store of texts translated from their ancient languages, RSTI serves as a tool for addressing research questions. To this end, the project also seeks to integrate archaeological data from the excavations at → Ras Šamra, including published archaeological plans, grid and square systems, and any other information freely available ... RSTI presents a standard text edition with translation, epigraphic commentary, philological explanations, and interpretation ... Objects, texts, and images are the three main categories of data in RSTI. The database has entries for 5,700 objects, mostly → tablets, but also vessels, seals, axe heads, and other → items. To date we have added 950 text → transliterations. The project currently includes over 32,000 tablet photos and drawings.

Miller C. Prosser, 314, 317

→ *item, Online Cultural and Historical Research Environment, Ras Šamra, research database environment, tablet, transliteration*

Raster Image

A matrix of numerical data for the computer called pixels.

Doğu Kaan Eraslan, 296

→ *Optical Character Recognition*

RDF → Resource Description Framework

RDF-XML

A syntax that can be used to express → Resource Description Framework (RDF) in an → Extensible Markup Language (XML) document.

Terhi Nurmikko-Fuller, 339

→ *Extensible Markup Language, Resource Description Framework*

Regular Expressions

A group of characters corresponding to a multitude of characters; thus, the characters of a regular expression are meta-characters. For example, in the Python implementation of regular expressions, the characters \w, \d correspond to an alphabetical character and a digit, respectively. There are various implementations of regular expressions throughout the programming languages.

Doğu Kaan Eraslan, 293n41

Relational Data Model

Similar → items are stored in tables of columns and rows, then joined with other tables based on a common column called a key.

Miller C. Prosser, 319

→ *item*

Relational Database

Computer database in which data are organized in tables, where each table contains all the instances of one entity. Each → “tuple” in a table represents one instance of the entity and must have a unique → identifier as one of its attributes; the columns (fields) of the table represent all other attributes of the entity, including keys of related entities. The relationships between the tables are defined as links between keys.

Sveta Matskevich and Ilan Sharon, 47n84

Relational databases are a type of data format that uses unique → identifiers (ID) to represent data from one table in another, so that when the user → queries the data using a query language called → Structured Query Language (SQL), it is possible to fetch related information from multiple tables at once.

Émilie Pagé-Perron, 196n7

→ *identifier, MySQL, querying, Structured Query Language, tuple*

Research Database Environment

It expands on the core structure of the database with tools and other features to help users work with their data.

Miller C. Prosser, 314n2

→ *Ras Shamra Tablet Inventory*

Resource Description Framework (RDF)

An abstract → data model with roots in → metadata modelling, it is used to represent information, knowledge, and data entities in a → graph structure. This approach captures the relationships between entities by representing these entities and the connections between them using → HyperText Transfer Protocol (HTTP) URIs ... As a flexible data structure, RDF is both powerful and capable of capturing nuanced and domain-specific relationships between data entities. Information stored in this way can also be restructured, manipulated, and edited with comparative ease without the extensive restructuring that a → relational database would require. RDF as a technology imposes few limitations on the type of information that can be captured and represented, whether at the level of entities (things, people, places, concepts, notions) or at the level of the relationships between them. All RDF data structures are based around a data cluster of three components, known as a → “triple.”

Terhi Nurmikko-Fuller, 339, 345

→ *data model, graph, HyperText Transfer Protocol Universal Resource Identifiers, meta-data, relational database, Simple Knowledge Organization System, triple*

RSTI → *Ras Shamra Tablet Inventory*

Scalable Vector Graphics (SVG)

→ Extensible Markup Language (XML)-based language for describing 2D graphics. SVG presents the most viable option for storing local glyph representations. SVGs are a joint product of the web development and design communities, making them reliable for expressing any type of 2D graphics that require cross-platform exchange. There are also different types of cross-platform open-source software for visualizing SVGs. The XML nature of SVGs implies a relatively simple integration with → EpiDoc.

Doğu Kaan Eraslan, 304

→ *EpiDoc, Extensible Markup Language, vector*

Scalar Multiplication

A multiplication operation between a → vector and a number that belongs to the same set of numbers contained in the vector.

Doğu Kaan Eraslan, 284n6

→ *vector*

Scraper

A type of software that selectively retrieves information from a website when a → machine-actionable version of the data is not available ... Nick Veldhuis designed a scraper specialized in extracting the → lemmata assigned to each → token of a text annotated in → Open Richly Annotated Cuneiform Corpus (Oracc). When used on an already annotated text, this type of tool will make it easier to gather specific types of entities: for instance, people in a text are → tagged with an indicator that the word is a personal name: “ [PN] ”.

Émilie Pagé-Perron, 204, 204n41

→ *annotation, lemma, machine-actionable/machine-readable, Open Richly Annotated Cuneiform Corpus, tag, token*

Self-Organizing Map (SOM)

Self-Organizing Maps are a type of → Artificial Neural Network (ANN); they have had great success and a number of concrete applications in many different fields. The logic underlying this type of ANN is the progressive distribution of data in an abstract map to form groups (potentially classes or sub-classes), which can show, based on their measurable distances, the degree of similarity that has been recognized among them.

Alessandro di Ludovico, 94n26

A type of learning → algorithm used to draw inferences from datasets and to order high-dimensional statistical data so that neighboring → nodes on the map represent similar inputs. Often the SOM is applied to numerical data in application areas such as pattern recognition, signal processing, and → multivariate statistical analysis. In other

words, the SOM is an → unsupervised type of network that offers a classification of the input vectors, creating a prototype of the classes and a projection of the prototypes on a map having two dimensions (but n -dimensional maps are also possible) that is able to record the relative proximity (or neighborhood) between the classes.

Marco Ramazzotti, 65n20

→ *algorithm, Artificial Neural Network, multivariate analysis, node, unsupervised method*

Semantics of a Recording System

In archaeological context, this is what is being recorded, irrespective of syntax. Only two types of archaeological entities are common across the whole spectrum of existing recording methods, systems, and schools. The two mandatory pieces of information that must be recorded by an excavator in every system are a find designator and the spatial unit from which that find originated. Many systems have more than one unit of either type.

Sveta Matskevich and Ilan Sharon, 38

→ *arbitrary unit*

Semantic Technologies

A heterogeneous collection of tools and → algorithms that can be used to bring structure to information. These include → Natural Language Processing, → data mining, → Artificial Intelligence, and a number of other approaches and methodologies.

Terhi Nurmikko-Fuller, 344

→ *algorithm, Artificial Intelligence, data mining, Natural Language Processing*

Semantic Web (sw) Technologies

A family of technologies and standards provided by the → w3c for the purposes of describing and connecting information online. A by-no-mean-exhaustive list of these technologies includes → Resource Description Framework (RDF), RDF Schema (RDFS), Web Ontology Language (OWL), and SPARQL.

Terhi Nurmikko-Fuller, 344

→ *Resource Description Framework, Simple Knowledge Organization System, web, World Wide Web Consortium*

Shading Attribute

It corresponds to a feature of the → Manuel de Codage (Mdc) encoding scheme: for example, A1\shading13, which would add diagonal lines to the first half of the → sign, indicating damage on that particular area.

Doğu Kaan Eraslan, 295n46

→ *Manuel de Codage, sign*

Sign → see General Glossary

Simple Knowledge Organization System (skos)

A model for expressing vocabularies, thesauri, and taxonomies in a → machine-readable format, namely → Resource Description Framework (RDF). SKOS is a → World Wide Web Consortium (w3c) standard employed for over a decade for linking data in the → Semantic Web. Controlled vocabularies and taxonomies are essential for efficient data retrieval in archaeological practice in general, and field recording in particular. Presented in SKOS format, they can be integrated into the → Linked Data cloud to serve all participant datasets.

Sveta Matskevich and Ilan Sharon, 46

→ *data cloud, Linked Data, machine-actionable/machine-readable, Resource Description Framework, Semantic Web, web, World Wide Web Consortium*

skos → Simple Knowledge Organization System

skosifying

A process of → mapping user-defined terminology (thesaurus or vocabulary) to → Simple Knowledge Organization System (SKOS).

Sveta Matskevich and Ilan Sharon, 48n91

→ *map/mapping, Simple Knowledge Organization System*

Social Network Analysis

Its main task is to explore social relations, and the → network graph data structure is the most appropriate data type to focus on relationships between entities ... The current practice in social network analysis based on cuneiform sources uses verbs to create directed links between entities or concepts.

Émilie Pagé-Perron, 195, 197

→ *administrative text, network graph*

som → Self-Organizing Map

SPAD → Système Portable pour l'Analyse des Données

Spatial Unit → see General Glossary

Special Characters

Exemplified by “*, _, =” as opposed to alphanumeric characters such as “a, b, 1, 2,” etc. ... Special characters do not solely express a particular semantic or syntactical field of the language. Some, for example, express phonetic values or diacritics ... Special

characters such as “_” or “\$” are reserved for indicating → sign functions and information about the state of conservation of the script and its → physical medium.

Doğu Kaan Eraslan, 285, 286n13, 288

→ *CAL Code, physical medium, sign*

SQL → *Structured Query Language*

Structured Query Language (SQL)

It is used to → query → relational databases to fetch data.

Emilie Pagé-Perron, 196n7

→ *MySQL, querying, relational database*

Sumerian → see General Glossary

Supervised Method

Method in which the researcher has chosen to “classify” part of the material before the analysis. The researcher is supervising, commonly called *training*, the → algorithm by providing it with known data prior to performing analysis on the unknown material.

M. Willis Monroe, 260, 260n9

→ *algorithm, unsupervised method*

SVG → Scalable Vector Graphics

Syntactic Elements

The components of the record itself, irrespective of what is being recorded. The most basic of these is the individual recording → event.

Sveta Matskevich and Ilan Sharon, 36

→ *event*

Système Portable pour l'Analyse des Données (SPAD)

A software for statistical analysis that has a → Graphical User Interface and includes a large series of different packages and functions. With the tools included in SPAD, one can perform different types of analyses and also manage data and whole datasets, with the aim, for example, of investigating their structure or preparing them to be processed. It is a proprietary software offering user interfaces in French. SPAD was first conceived as academic software, but after 1987 it became a commercial product developed by CISIA/DECISIA.

Alessandro di Ludovico, 97n35

→ *data mining, Graphical User Interface*

Tablet → see General Glossary

Tag

When using → Extensible Markup Language (XML) or → Text Encoding Initiative (TEI), it is rather an indication of the classification of what is described: after a tag “word,” one either has a verb, adverb, noun, or adjective, under its inflectional form or not.

Vanessa Bigot Juloux, 163

In → EpiDoc, the appellation of tags filling the role of → markups, as in the case of → Extensible Markup Language (XML) documents, must be decided in the process of developing the → namespace. Deciding what a tag should be called, however, does not necessarily result in its coherent usage. Taking the example of the name John Doe Smith, we would not know whether the name belongs to a real person or a fictional character. If we were to distinguish this aspect of the names in a document, the distinction would not come from the domain of the namespace, since it is possible for the same name to belong to both real and fictional persons. Rather, the differentiation would need to be made through the manner in which we apply the tags. Therefore, it is essential that tag usage is a planned part of a project.

Doğu Kaan Eraslan, 302n72

→ *attribute, EpiDoc, Extensible Markup Language, markup/markup tagging, namespace, Text Encoding Initiative*

TEI → Text Encoding Initiative

Term Frequency–Inverse Document Frequency (TF-IDF)

A quite popular method that evaluates the importance of terms within one text inverse to their frequency in the entire corpus, essentially identifying the terms that are most important for identifying a document. This method is based on frequency counts for terms within documents and within the corpus as a whole. It works well when there are many terms shared among all documents, as it weeds out common terms and focusing on the important rare terms.

M. Willis Monroe, 271–272n34

Text Data

A glyph, a word, a group of words, a reference, or a concept.

Vanessa Bigot Juloux, 165

Text Encoding Initiative (TEI)

TEI was created in 1987. Then, in 2000, the TEI Consortium (TEI-C) was established. The TEI-C is a group of international scholars who collaborate on the development of a

dedicated encoding standard for text analysis. Unfortunately, using TEI is a “paradox,” since it does not yet enable → interoperability, but rather the interchange of cross-corpora text analysis.

Vanessa Bigot Juloux, 164

→ *attribute, element, EpiDoc, interoperability/epigraphic interoperability*

Text Mining

Its purpose is to find patterns that match shared analytical variables in order to suggest a first interpretation of the transcription.

Vanessa Bigot Juloux, 181

→ *qualitative method*

TF-IDF → Term Frequency–Inverse Document Frequency

Token

A sentence or → sign, but, in this case, the texts are divided at the word level. Tokens are then assigned to → lemmata, which are also called “dictionary entries,” in a process called “lemmatization.”

Émilie Pagé-Perron, 203

→ *lemma, sign*

Tokenization

The tokenization process consists of splitting the group of words into individual words. Each word is sorted into a category, the equivalent of a → Text Encoding Initiative (TEI) → attribute value.

Vanessa Bigot Juloux, 166, 203

It is the segmentation of a text into similar units called → tokens ... in a process by which occurrences of a word, or other units (such as → signs or sentences), are separated as units from the original textual data. Working from → transliterations, the major separator to take into account is of course the space character.

Émilie Pagé-Perron, 204n41

→ *attribute, sign, Text Encoding Initiative, token, transliteration*

Topological Weighted Centroid (TWC)

A topological approach with a set of ordered mathematical quantities that transforms the discrete dataset into different scalar fields: Alpha, Beta, Gamma, Theta, and Iota.

Paolo Massimo Buscema, in Marco Ramazzotti, 74

Topological Weighted Centroid (twc) Mathematical Approach

It explores the natural and anthropic landscape with respect to certain quantities (entropy) spatializing the optimal solutions in the centers of the masses.

Marco Ramazzotti, 70n30

→ *Topological Weighted Centroid*

Transliteration → see General Glossary

Triple

It has two distinct meanings:

1. Triples consist of three parts: a SUBJECT, a PREDICATE, and an OBJECT. These are used to define entity types and directional relationships between them, both at the general (or schema) and specific (instance) level.

Terhi Nurmikko-Fuller, 345

2. It always has two → nodes and an → edge (individual 1, individual 2, and the relationship between them).

Émilie Pagé-Pérro, 202

→ *edge, network graph triples, node, triple store*

Triple Store

A database for the storage and retrieval of → triples.

Sveta Matskevich and Ilan Sharon, 46n82

→ *triple*

Tuple

An ordered list or row in a table.

Sveta Matskevich and Ilan Sharon, 47n84

twc → Topological Weighted Centroid

twc Mathematical Approach → Topological Weighted Centroid Mathematical Approach

Unicode

A system to map characters to bytes through code points. Characters are treated as semantic units, and code points are arbitrary constructions. The → mapping requires two different reversible procedures that make the system rather complex. Most modern platforms use a Unicode-based encoding scheme, since it was designed to support every character used in every language.

Doğu Kaan Eraslan, 300–301

→ *map/mapping, Unicode Consortium, Unicode Transformation Format-8*

Unicode Consortium

Computers must use an agreed-upon system of representing characters. The Unicode Consortium was created to promote this standard. One of the most significant contributions of this group is the creation of a standard that defines which underlying computer code is used to represent (nearly) every character in every writing system, even → Ugaritic. The number of each character can be referred to as a → Unicode point.

Miller C. Prosser, 328n34

→ *Ugaritic, Unicode*

Unicode Transformation Format (UTF)-8

“8” refers to the use of 8-bit (a numerical value that equals 0 or 1) sequences to represent a character. It is a method for encoding → Unicode characters. There is also UTF-16 and UTF-32, which have, respectively 16- and 32-bit character sequences.

Doğu Kaan Eraslan, 289n25

→ *EpiDoc, Unicode*

Universal Resource Identifiers (URIs)

These clusters of letters and characters form unique → identifiers for each and every page on the → web ... On a conceptual level, these URIs do not differ from any other unique identifier, such as an ISBN for a book, a national identity number for a person, or an object identifier in a museum, but they are unique on the scale of the entire web ... When utilized consistently and documented appropriately, URIs can and are used by large numbers of data publishers (people and institutions who make data available online) to show that completely disparate and unconnected pages contain information about the same entity or resource.

Terhi Nurmikko-Fuller, 338–339

→ *HyperText Transfer Protocol, identifier, HyperText Transfer Protocol Universal Resource Identifiers, web*

Unsupervised Method

Classification method that involves not classifying any of the existing corpus, but, rather, letting the → algorithm decide which records out of the entire corpus belonged together ... The benefit of the unsupervised method here, however, is that it breaks what are otherwise researcher-oriented biases and presents an “objective” view of the material, free from prior classification. Perhaps the unsupervised method detected that a group of novels thought to belong to one author most closely match the unknown material in our example. The conclusion is up to the researcher here, but the unsupervised method can offer insight into a dataset without any preconceived notions of structure and definition. Unsupervised methods are commonly used to explore a dataset, after which a → supervised method can be used to hone in on a particular area of interest.

M. Willis Monroe, 260–261

→ *algorithm, Self-Organizing Map, supervised method*

Unweighted Pair Group Method with Arithmetic Mean (UPGMA)

A → clustering method that is relatively simple in its procedure. It starts by looking for the two closest cells within the adjacency matrix and merging those into a cluster to which it assigns the average of their distance. It then looks for the next two closest cells, or, if the next cell is closest to the newly created cluster, it merges those and assigns it a new average distance. This process is completed for the entire adjacency matrix, merging close items whether they are cells or newly created clusters.

M. Willis Monroe, 275

→ *clustering method*

UPGMA → Unweighted Pair Group Method with Arithmetic Mean

Upper Ontology

A highly generic, non-specifying schema of data categories applicable for use across many research domains. The categories of data types, such as Location, Person, Text, and Resource, are very broad. Each category of data is slightly different from every other, both in conceptual definition and in practical implementation. These data types are presented to the user as different hierarchies.

Miller C. Prosser, 320

→ *ontology*

URIs → Universal Resource Identifiers

Vector

Mathematically speaking, for example in the context of → Scalable Vector Graphics (svg) and → Optical Character Recognition (OCR), a vector is anything that represents something that has a size and a direction. Size is the distance between the origin and end points, and direction is represented by the angle that a line makes with an axis.

Doğu Kaan Eraslan, 284n6

→ *Optical Character Recognition, Scalable Vector Graphics, scalar multiplication, vector addition, vector graphic, vector space*

Vector Addition

It simply entails adding two or more → vectors together.

Doğu Kaan Eraslan, 284n6

→ *vector*

Vector Graphic

A group of lines passing through points defined in a coordinate system.

Doğu Kaan Eraslan, 297

→ *Optical Character Recognition, vector*

Vector Space

In linear algebra, an additive group that is associated with a field of real numbers and permits → vector addition and → scalar multiplication.

Doğu Kaan Eraslan, 284n6

→ *scalar multiplication, vector, vector addition*

W₃C → World Wide Web Consortium

Web

A hypermedia system of pages, sites, and other types of resources, accessed and connected through the use of → hyperlinks ... Although its current manifestation is one of interconnected pages, the three technologies that lie at the heart of the web—→ HyperText Markup Language (HTML), → HyperText Transfer Protocol (HTTP), and → Universal Resource Identifiers (URIs)—are also at the core of → Linked Data (LD) ... The web currently consists of over 4 billion pages, all connected internally (within the page and a site) and externally (to other pages on other sites) through hyperlinks.

Terhi Nurmikko-Fuller, 338

→ *hyperlinks, HyperText Markup Language, HyperText Transfer Protocol, Linked Data, Universal Resource Identifiers, World Wide Web Consortium*

Weight

The weight of an → edge represents here the quantity of co-occurrences of two individuals in the corpus.

Émilie Pagé-Perron, 201

→ *edge, weighted*

Weighted

Weighted relationships can be highlighted in different ways. In → Gephi, one can easily make the edges thicker based on the numeric value of their → weight. In → Cytoscape, a practical way to display the frequency or weight of an → edge is either by applying a color range or by changing the thickness of the edge, with the variation depending on the weight value.

Émilie Pagé-Perron, 211n61

→ *Cytoscape, edge, Gephi, weight*

Wizard

Wizards aid us in a variety of tasks that help us address a research question. There are currently three wizards for philological analysis: (1) the → lexicography wizard; (2) the prosopography analysis wizard; and (3) the gazetteer wizard.

Miller C. Prosser, 329

→ *lexicography, workflow wizard*

Word Sense Induction

Also called word sense discrimination or word sense discovery, it is the task of determining what meaning a word may have in different contexts.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 226–227

Word2vec

Developed by Tomas Mikolov and his team in 2013 ... Word2vec is an open-source toolkit for producing word vectors and → querying semantic relationships between words ... It uses so-called → Artificial Neural Networks to predict word co-occurrence ... There are two alternative models one can choose from when using Word2vec: the → Continuous Skip-gram model and the → Continuous Bag-of-Words (CBOW) model.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 229, 246–247, 247n57 → Artificial Neural Network, Continuous Bag-of-Words model, Continuous Skip-gram model, Natural Language Processing, Pointwise Mutual Information, querying

Workflow Wizard

An interactive database tool that guides the user through a series of common actions.

Miller C. Prosser, 314n3

→ *wizard*

World Wide Web Consortium (w3c)

w3c is an acronym referring to the World Wide Web Consortium, the international standards organization for the World Wide Web.

Terhi Nurmikko-Fuller, 340n7

→ *web*

Writing System

In → Online Cultural and Historical Research Environment (OCHRE) we have created a writing system that represents the → logosyllabic → cuneiform writing system used in the ancient Near East ... Like the list of letters that represents the English alphabet, this list represents all the known → signs from ancient languages such as → Sumerian and → Akkadian.

Miller C. Prosser, 321

→ *Akkadian, cuneiform, logosyllabic, Online Cultural and Historical Research Environment, sign, Sumerian*

XDR → External Data Representation

XML → Extensible Markup Language

XML Annotation → Extensible Markup Language annotation

XML-Elamite Standard

Developed by Marc Bavant, it is tailored for conserving → Elamite texts along with their → transliterations and translations. It is based on → Unicode characters, and it follows widely used conventions in Elamite scholarship, making it → human readable.

Doğu Kaan Eraslan, 287

→ *Elamite, human readable, transliteration, Unicode*

General*

(including terms associated with archaeology, history, geography, literature, and philology)

Adab Corpus

It dates to the third millennium BCE and comprises all texts from Adab. The texts include both those from official excavations and those that are lacking provenance but have been attributed to Adab based on factors such as the shape of the → tablet, prosopography, and the institutions and rulers mentioned. The large majority of the texts are → administrative in nature.

Émilie Pagé-Perron, 210n60

→ *administrative text, tablet*

Administrative Texts

They mostly document transactions involving people, things, actions, and places ... Administrative texts are typically short and → formulaic, with simple sentences and few actions ... Despite comprising the most numerous type of surviving → cuneiform documents, administrative texts are the least-annotated genre of → Mesopotamian sources ... Because they also remain untranslated, they have the additional drawback of being inaccessible to specialists in adjacent disciplines who are not trained to read cuneiform.

Émilie Pagé-Perron, 195

→ *Adab corpus, annotation, cuneiform, formulaic, Mesopotamia*

* Although all words of an expression are capitalized in the list of terms below, when indicated after the “see also” arrow, words are capitalized when the expression has an acronym (for example, Electronic Distance Measurement, EDM) but not capitalized when the expression does not have an acronym. The “definitions” provided in this glossary are neither exact nor comprehensive; rather, they are informative statements extracted from the chapters. In some cases, minor modifications (such as verb tense, or the addition of “it is”) have been made in order for the statement to make sense out of context. When multiple statements are presented for a term, they are listed according to the alphabetical order of their authors’ names. Cross-references point both to entries in this glossary as well as in the CyberResearch Glossary.

Akkadian

The Akkadian → writing system allowed for variation in the spelling of a given word, even among words with the same grammatical properties.

Miller C. Prosser, 322

A Semitic language in use from about 2500 until 500 BCE ... The script uses both → syllabic and logographic signs, and each → sign can be read in many different ways. Akkadian also uses complex inflection; that is, words are modified in order to express various grammatical categories ... For example, *šarrum*, “king,” for the singular nominative form of the noun and *šarram* for the singular accusative form.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 224n3, 227
→ *logosyllabic, sign, syllabic sign, writing system*

Allograph

A variant form of the letter, exemplified in most modern alphabetic systems by uppercase and lowercase letters.

Miller C. Prosser, 321n26

→ *script unit*

Annotation

In corpus linguistics, an annotation is a comment that specifies the various linguistic features of a word.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 225n8

Aramaic

An ancient Northwest Semitic language attested throughout eastern Turkey, northwest Iran, Iraq, Syria, the Levant, and parts of Arabia at the beginning of the 1st millennium BCE and onwards. Its descendants, such as Mandaic, are still spoken today.

Doğu Kaan Eraslan, 288n20

Arbitrary Unit → see CyberResearch Glossary

Archaeological Site

An accumulation of archaeological remains spatially related to each other.

Sveta Matskevich and Ilan Sharon, 48

Area

A subdivision of the → archaeological site (usually several contiguous grid squares, encompassing at least one, possibly several, architectural features).

Sveta Matskevich and Ilan Sharon, 48

→ *archaeological site*

Attested Form

A form of the → lemma as it occurs in a text, represented by a sequence of → signs.

Miller C. Prosser, 322

→ *lemma, sign*

Attribute (for digital practice) → see CyberResearch Glossary

Attribute (in Archaeology)

It is a minimal characteristic of an artifact such that it cannot be further subdivided.

Sveta Matskevich and Ilan Sharon, 33n37

Ba'lu and 'Anatu Cycle (KTU 1.1–6)

A narrative story, traditionally viewed as a “myth” by scholars, about the fight between two clans over the throne. Written in alphabetic → cuneiform on six double-sided clay → tablets, the text is attributed to the scribe 'Ilimilku the Šubbanite of → Ugarit.

Vanessa Bigot Juloux, 152

→ *cuneiform, tablet, Ugarit*

Basic Volume Unit (BVU)

In archaeological context, the primary excavation unit of a recording system, defined arbitrarily (grid square), stratigraphically (→ locus, context), or behaviorally (feature).

Sveta Matskevich and Ilan Sharon, 49n93

→ *locus*

BVU → Basic Volume Unit

Colophon

A brief statement on a text, such as a → cuneiform → tablet, that captures the equivalent of what could now be considered bibliographical → metadata and can contain information such as the name of the authoring individual.

Terhi Nurmikko-Fuller, 345n23

→ *cuneiform, metadata, tablet*

Composite Texts

Composites may represent the content of a single instance of a → tablet of any degree of preservation or completeness, or of any number of witness tablets, numbering in some cases in the hundreds.

Terhi Nurmikko-Fuller, 342

→ *tablet*

Cuneiform

It is widely accepted that cuneiform writing was invented to express the → Sumerian language. The → writing system was later adopted by other cultures to express their own languages.

Miller C. Prosser, 316n13

The cuneiform script was originally developed by Sumerians c. 3000 BCE. In the following centuries, the script was adapted for → Akkadian. It was also used to write several other ancient Near East languages, such as → Hittite and → Elamite. Although passing centuries and new languages caused the form of the → signs to change, the basic shapes—triangular wedges left by a reed stylus—stayed the same.

Saana Svärd, Heidi Jauhiainen, Aleks Sahala, Krister Lindén, 224n3
→ *Akkadian, Elamite, Hittite, sign, Sumerian, tablet, Ugaritic, writing system*

Daily Top-plan

Daily graphical record of an excavation area or part of it. The background plan shows all architectural features extant in the area on that particular day. The foreground information reflects how the excavation proceeded during the day. It comprises several types of data, such as outlines of excavation units, elevations, and the registration numbers and coordinates of finds.

Sveta Matskevich and Ilan Sharon, 26n10

Demircihüyük

Site located in northwest Turkey near modern-day Eskişehir. It has remains from the Neolithic to the second millennium BCE.

Shannon Martino and Matthew Martino, 124

Dor (Tel Dor)

A major port town on the Mediterranean, between present-day Tel Aviv and Haifa in Israel. It was active from the Middle Bronze Age to the Roman era.

Sveta Matskevich and Ilan Sharon, 25

Ecofact

A non-artifactual organic (botanical or biological) or inorganic (result of geological processes) object recovered from an archaeological context.

Sveta Matskevich and Ilan Sharon, 42n68
→ *archaeological site*

EDM → Electronic Distance Measurement

Elam

Ancient civilization existing from the third into the mid-first millennium BCE based in the southwest region of the Iranian plateau, just northeast of the Persian Gulf (in the modern provinces of Ilam and Khuzestan).

Doğu Kaan Eraslan, 287n16

Elamite

An ancient language used in → Elam.

Doğu Kaan Eraslan, 287n16

→ *cuneiform, Elam*

Electronic Distance Measurement (EDM)

An instrument that uses reflected light (infrared, laser, microwave) to calculate distances to remote objects; the models produced in the 1970s through the 1990s were mounted on top and externally connected to a theodolite. Newer models, in which a digital theodolite and an EDM are integrated within the same instrument, are known as “Total Stations.”

Sveta Matskevich and Ilan Sharon, 27n15

Formulaic

Administrative texts of the → Old Akkadian period are formulaic in the sense that they mostly consist of lists of people, things, or both, with verbs that have specific, technical meanings in particular contexts. There is little deviation from the usual formula.

Émilie Pagé-Perron, 201n32

→ *Old Akkadian period*

Glyph

A graphical representation, such as a character or an accent. In → Ugaritic → cuneiform script, it is one or several carved lines or impressed wedges expressing a letter, syllable, logogram, or a separation mark (a dot) between two words.

Vanessa Bigot Juloux, 165n69

The most basic part of the visual representation of an elementary semantic unit. In the case of an Egyptian hieroglyph, for example, a glyph could be a sitting man or a bird. For → cuneiform, a glyph would be a horizontal, vertical, or diagonal line comprising a cuneiform → sign, and for alphabetic languages, such as Greek and Aramaic, a glyph corresponds to a visual representation of a letter, such as an alpha.

Doğu Kaan Eraslan, 291n32

→ *cuneiform, sign, Ugaritic*

Grapheme

→ Syllabic → sign or the alphabetic letter.

Miller C. Prosser, 323

→ *sign, syllabic sign*

Hittite

An Indo-European language of Anatolia (Asia Minor) written with → cuneiform → signs.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 227n11

→ *cuneiform, sign*

Intermediary Elements

Vehicle between what we see in the → physical medium and what we understand. This corresponds, for example, to the alpha-numeric characters that are employed to represent the → glyph in the physical medium.

Doğu Kaan Eraslan, 291n31

→ *glyph, physical medium*

KTU 1.1–6 → Ba'lu and 'Anatu Cycle

Layer

In Archaeology, a unit of sediments in a stratified → archaeological site, created as a result of one of the site-formation cycles.

Sveta Matskevich and Ilan Sharon, 42n67

→ *archaeological site, stratum, tell*

Lemma (pl. lemmata)

A lemma is the headword used in a dictionary entry.

Émilie Pagé-Perron, 203n38

In → OCHRE, a lemma is further defined by → phonemic forms, which in turn are defined by → attested forms.

Miller C. Prosser, 321–322

→ *attested form, Online Cultural and Historical Research Environment, phonemic form*

Lexicography

The practice of compiling glossaries and dictionaries.

Miller C. Prosser, 329

Locus (pl. loci)

The basic → spatial unit in most recording systems in Near Eastern Archaeology. The exact denotation of this term (i.e., the archaeological entity that is modeled by the abstract unit) can vary among different systems ... The de facto standard definition of the “locus” in Levantine Archaeology today on sites with architecture is a contiguous segment of the site volume interpreted as being the result of a single depositional event.

Sveta Matskevich and Ilan Sharon, 26n9, 40

→ *spatial unit*

Logosyllabic

A → writing system that uses characters—or → “signs”—that can represent words or syllables. For the → Akkadian logosyllabic writing system, we call these word-signs logograms and the syllable-signs phonograms. Thus this type of writing is called logosyllabic, a combination of logograms and syllables.

Miller C. Prosser, 316n11

→ *Akkadian, sign, syllabic sign, writing system*

Mesopotamia

A region situated within the Tigris–Euphrates river system, in modern days roughly corresponding to most of Iraq and the eastern parts of Syria. A number of ancient cultures and political units flourished in the area.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 224n2

→ *Mesopotamian civilizations*

Mesopotamian Civilizations

Mesopotamian civilizations are known through their texts, mostly written in → Akkadian and → Sumerian, as well as through archaeological finds from the region.

Saana Svärd, Heidi Jauhiainen, Aleksi Sahala, Krister Lindén, 226

→ *Akkadian, Mesopotamia, Sumerian*

Mesopotamian Urban Revolution Landscape (MURL)

Historical period and economic concept of the Urban Revolution, as well as the area of the world's first urbanism—central-southern Iraq.

Marco Ramazzotti, 62n10

→ *Mesopotamia, Mesopotamian civilizations*

Metadata → see CyberResearch Glossary

Micro-zodiac Text

The Micro-zodiac material comes from Hellenistic Babylonia, where it was written by scribes educated and working in the Late Babylonian scholarly communities of southern → Mesopotamia ... The corpus of the Micro-zodiac texts consists of 15 known → tablets of varying degrees of preservation. The most complete have more than 20 cells (the unit of analysis for this paper) of ingredients on them.

M. Willis Monroe, 259n4, 261

→ *Mesopotamia, Mesopotamian civilizations, tablet*

MURL → Mesopotamian Urban Revolution Landscape

Neolithic Package

A collection of technologies and beliefs that was once thought to be wholeheartedly and without exception adopted along with agriculture.

Shannon Martino and Matthew Martino, 115

Old Akkadian Period

Approximately from 2340 BCE to 2200 BCE: a period in → Mesopotamian history during which this region saw the development of an important political entity that, at its apogee, extended almost from the Mediterranean in the west, to the Persian Gulf in the south, and to the Zagros mountains in the east.

Émilie Pagé-Perron, 196n10

→ *Mesopotamia, Mesopotamian civilizations*

Old Persian

An Indo-Iranian language descended from Old Iranian. It was used mostly, but not exclusively, in royal inscriptions in the Achaemenid Empire (550–330 BCE).

Doğu Kaan Eraslan, 301n70

Online Cultural and Historical Research Environment (OCHRE) → see CyberResearch Glossary

Phonemic Form

It represents a grammatical interpretation of an → attested form. In many cases, a given phonemic form is represented by many attested forms.

Miller C. Prosser, 322

→ *attested form*

Physical Medium

An object on which we observe the → signs. It can be papyrus, a vase, a stone surface, etc.

Doğu Kaan Eraslan, 285n8

→ *sign*

Ras Šamra

The → archaeological site of Ras Šamra is located on the eastern Mediterranean, near Lattakia, Syria. The ancient name for this site—and for the surrounding kingdom—was → Ugarit. Ugarit was well situated, with access to trade routes, both land and sea, and to arable lands. The site was occupied almost continuously from the Neolithic period (c. 7000 BCE) through its eventual destruction in the twelfth century BCE during a regional period of instability. Later, Greek, Persian, and Roman garrisons occupied the site.

Miller C. Prosser, 315

→ *archaeological site, Ugarit*

Raw Data of an Excavation

A collection of interpretations fixed in pre-printed forms, diaries, and tables.

Sveta Matskevich and Ilan Sharon, 44

Reading

A value represented by a → sign [as part of a → script unit].

Miller C. Prosser, 321n26

→ *script unit, sign*

Script Unit

It is more than just a letter or a → sign in the → writing system. Each script unit is defined by various → readings and → allographs.

Miller C. Prosser, 321

→ *allograph, reading, sign, writing system*

Sign

Elementary semantic unit comprising alphabetical and non-alphabetical languages.

Doğu Kaan Eraslan, 285n8

Spatial Unit

A locational unit of excavation defined in absolute coordinates, relatively to the site grid, or to architectural features and other units in its vicinity.

Sveta Matskevich and Ilan Sharon, 26n9

→ *arbitrary unit*

Static Archaeological Record

One of the keywords of the middle-range theory in Archaeology; it refers to all findings and the fact that, although uncovered almost simultaneously and in close physical proximity, they represent some diachronic reality.

Sveta Matskevich and Ilan Sharon, 31n26

Stratum

A single construction–destruction cycle within a stratified sequence of deposits that form a multi-layered site (e.g., → tell).

Sveta Matskevich and Ilan Sharon, 40n61

→ *archaeological site, layer, tell*

Sumerian

In the context of literary Sumerology, “Sumerian” is to be understood as an exclusively linguistic label, referring to the language of the compositions themselves, rather than being indicative of the ethnic or socio-cultural identity of the authoring scribe or the patron commissioning the piece. Data is not limited to sources with a proven provenance in Sumerian periods for two reasons: First, literary pieces unearthed from third-millennium BCE contexts at geographically diverse sites (Ebla and Mari in the north and Girsu, Abu Salabiḥ, Nippur, and Adab in the south) suggest these compositions were conceived of even earlier; second, while primary sources for Sumerian literature originate from a range of temporal, geographical, and cultural settings, many are dated to the Old Babylonian (OB) period (c. 20th–16th centuries BCE). Sumerian is absent from → administrative and legal documents, as well as letters, from around 1730 BCE onwards, supporting the theory that at this time it no longer survived as a spoken language in any of the urban centers of the → Mesopotamian plateau. Its use in the vernacular registers of personal correspondences had ceased some two centuries earlier. Scribes who most likely spoke → Akkadian as their mother tongue carried on the tradition of Sumerian literature via copies of earlier compositions (e.g., Instructions of Šuruppak) as well as new compositions and those created through extensive modification of known material such as *Lugalbanda*.

Terhi Nurmikko-Fuller, 350–351

→ *administrative texts, Akkadian, cuneiform, Mesopotamia, Mesopotamian civilizations*

Syllabic Sign

It represents a syllable, whereas a logographic → sign represents a word.

Saana Svärd, Heidi Jauhiainen, Aleksī Sahala, Krister Lindén, 225n9

→ *cuneiform, logosyllabic, sign*

Syntactic Influence

It refers to the observation of the order of linguistic elements in a sentence of one language in a sentence of another, linguistically unrelated, language. This is exemplified by the formula “x-š-a-y-θ-i-y : x-š-a-y-θ-i-y-a-n-a-m,” “king of kings” in → Old Persian royal inscriptions. Normally in Old Persian, the genitive precedes its object *(“p-a-r-s-h-y-a : p-u-ç” [Xph line 12], son [p-u-ç] of Persian [p-a-r-s-h-y-a])* , but in → Akkadian the genitive follows its object, as in “šarrū māti,” meaning “kings of the land.” Hence, with the genitive “x-š-a-y-θ-i-y-a-n-a-m” following its object “x-š-a-y-θ-i-y,” as in the Akkadian example, the occurrence of the genitive “māti” after its object “šarrū” represents syntactic influence.

Doğu Kaan Eraslan, 301n70

→ *Akkadian, Old Persian*

Tablet

A clay object, probably one that fits comfortably in the hand and is inscribed with the distinctive wedge-like → signs of the → cuneiform script.

Terhi Nurmikko-Fuller, 343

→ *cuneiform, sign*

Tell

An artificial mound formed as a result of successive cycles of construction, occupation, and destruction, sometimes separated by periods of abandonment.

Sveta Matskevich and Ilan Sharon, 25n3

→ *archaeological site, layer, stratum*

Transliteration

Conversions of the → cuneiform script into the Latin alphabet without the translation of the ancient language to a modern one.

Terhi Nurmikko-Fuller, 341n11

Transcription of cuneiform inscriptions into a romanized rendering of the → cuneiform → sign readings. This includes marking word boundaries and some structural markers, such as line numbering, object surfaces, etc.

Émilie Pagé-Perron, 199n18

→ *cuneiform, sign*

Ubaid Clay Sickles

The Ubaid clay sickles (fifth to fourth millennium BCE) are usually associated with the first intensive agricultural activities in the alluvial plain.

Marco Ramazzotti, 67n29

→ *Mesopotamia, Mesopotamian civilizations*

Ugarit

The Kingdom of Ugarit had about two hundred villages, with an estimated population of thirty-five thousand inhabitants, including the capital. The kingdom was destroyed at the end of the Late Bronze Age, c. 1200 BCE.

Vanessa Bigot Juloux, 152n6

In the centuries prior to its final conflagration, Ugarit was a cosmopolitan culture, exhibiting artistic and stylistic influences from Egypt, → Mesopotamia, the Aegean, and Anatolia. To date, Ugarit has yielded approximately 4,500 texts in various languages and writing systems.

Miller C. Prosser, 316

→ *Ba'lu and 'Anatu Cycle, Mesopotamia, Ras Šamra, Ugaritic*

Ugaritic

The scribes of → Ugarit used a newly invented alphabetic writing system consisting of 30 letters ... Like the → logosyllabic Mesopotamian writing system, Ugaritic letters are formed by impressing a stylus into a clay → tablet, creating a series of wedges. This type of writing is known as → cuneiform, from the Latin *cuneus*, “wedge” ... The Ugaritic alphabetic → writing system indicates vowels only partially and only indirectly by use of three aleph → signs ʾ, i, and ʰ. Otherwise, no vowels were written. The vocalized form of the word is meant to convey the grammatical interpretation of the word.

Miller C. Prosser, 316, 325n32

→ *cuneiform, logosyllabic, sign, tablet, Ugarit, writing system*

Uruk Clay Cones

Uruk clay cones (fourth millennium BCE) are linked to the ideographical and symbolical representations of the most important religious buildings of the first southern → Mesopotamian cities.

Marco Ramazzotti, 67n29

→ *Mesopotamia, Mesopotamian civilizations*

Writing System → see CyberResearch Glossary