A burning question: using an intelligent grey literature search engine to change our views on early medieval burial practices in the Netherlands

Brandsen, A.; Lippok, F.E.

Journal of Archaeological Science 133 (2021) 105456

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: www.elsevier.com/locate/jas

A burning question – Using an intelligent grey literature search engine to change our views on early medieval burial practices in the Netherlands

Alex Brandsen *, Femke Lippok

Leiden University, the Netherlands

ARTICLE INFO

Keywords:
Grey literature
Document search
Early Middle Ages
Cremations
Natural language processing
Machine learning
Deep learning

ABSTRACT

This paper presents a case study on Early Medieval burial practices using AGNES, an intelligent search engine for Dutch archaeological grey literature. Traditionally, it is assumed that cremations phased out at the start of the Early Middle Ages, when the inhumation practice became more numerous. However, recent research (Lippok, 2020) shows that cremations might be more prevalent than previously assumed. Due to research efforts being concentrated on furnished inhumations, cremations and other types of burials have not received their share of research interest. It is suspected that unknown Early Medieval cremations may be found in grey literature research reports. The rapidly growing document collection requires more efficient methods to search through this big data, as manual searching is too time intensive.

AGNES uses machine learning to allow searching on archaeological concepts (such as time periods and artefacts) in full texts, solving common problems with synonymy and polysemy. This paper describes a controlled search for Early Medieval cremations in the Netherlands, and a comparison of the new information uncovered from the reports to prior knowledge on the topic. The queries resulted in 2541 hits. Twenty-three Early Medieval cremations that were previously unknown to experts were uncovered, and 31 (of 77) known sites were identified. Forty-one possibly interesting documents were noted, where it was not clear from the report if Early Medieval cremations were present. 2446 documents were not relevant to the study for various reasons.

The 23 new sites are an increase of 30% over the existing knowledge of experts. In the last 20 years only nine new sites featuring Early Medieval cremations were discovered, so being able to add another 23 is a major development. Adding previously unknown Early Medieval cremations to the dataset challenges the existing bias for inhumation graves, and supports striving towards viewing the Early Medieval burial repertoire as more heterogeneous and representative. This indicates that AGNES is useful for archaeological research, and the uncovered information can lead to a more cohesive view of the past.

1. Introduction

Archaeologists produce large amounts of texts, from monographs and articles to research reports written by commercial units. In the Netherlands, it is estimated that around 60,000 of these reports were produced up until 2017, with 4000 being added every year (RCE, 2019). Often, these reports are created in-house and not circulated widely, made available only via online repositories such as the DANS (Data Archiving and Networked Services) archive or the Dutch national cultural heritage database, Archis, maintained by the RCE (*Rijksdienst Cultureel Erfgoed*, the Dutch heritage agency).

Such reports are also known as grey literature: documents produced outside of the traditional commercial and academic publishers, often with a small audience and not peer-reviewed. These reports are currently underused especially for synthesising research, even though many authors note that the information in these documents can be of great value (Evans 2015; Richards et al. 2015; Brandsen et al. 2019).

For Dutch grey literature, it was until now only possible to search through the metadata of the documents, not the texts themselves, using the above mentioned DANS and Archis search systems. This is not ideal, as archaeologists will often want to search more fine grained, and might be interested in a single Iron Age artefact on an otherwise Medieval site, which is not mentioned in the temporal metadata as it is too specific. Such examples are often called 'by-catch', one or a few finds that are different from the rest of the excavation. This by-catch is currently impossible to find effectively, and archaeologists report they currently

* Corresponding author. Faculty of Archaeology, Einsteinweg 2, 2333CC, Leiden, the Netherlands.
E-mail addresses: a.brandsen@arch.leidenuniv.nl (A. Brandsen), f.e.lippok@arch.leidenuniv.nl (F. Lippok).

https://doi.org/10.1016/j.jas.2021.105456
Received 16 June 2021; Received in revised form 27 July 2021; Accepted 28 July 2021
Available online 3 August 2021

download large numbers of reports and manually search through each PDF file, a time consuming and inaccurate task (Brandsen et al. 2019).

One way to make this document collection more useful is by applying full text search. Similar to e.g. Google, this allows archaeologists to search through all of the text of all of the documents. This would already be an improvement to the current situation, but from previous research (Vlachidis et al. 2017; Habermehl 2019) and our own prior work (Brandsen et al. 2020) it is evident that for archaeological texts, synonymy and polysemy are common and can cause problems.

Synonyms are different words that have the same (or similar) semantic meaning. An example is the Middle Ages, which can also be written as the Medieval Period, the Dark Ages, 500–1500 CE, and includes subperiods such as the Merovingian Period, and dates such as 600 CE or 1050 ± 23 BP. Ideally, an archaeologist will want to find all of these mentions when looking for reports about the Middle Ages.

Polysemy is the other way around: one word having multiple meanings. An example is "Mayen", which can indicate a type of pottery, a pottery production site spanning ten centuries, a volcanic island in the Norwegian sea called "Jan Mayen" or a town in Germany (which the first two are named after). But when you type "Mayen" into a standard text search engine, the system does not know which meaning you are interested in, and will provide results about all the meanings, or maybe just the most popular meaning.

A combination of full text search with entity search is used to solve the potential problem of polysemy. We apply Named Entity Recognition (NER), a method which automatically finds and labels relevant concepts for archaeologists. This can be done using a rule-based approach, or — in the case of our study — by using Machine Learning. The NER process uses the context of words to attempt to distinguish between meanings. This process is not 100% accurate, as ambiguous entities are a challenge: e.g. "*aardewerk*" which can mean the artefact 'pottery', or the material 'ceramic' in Dutch. But where the NER is correct, users searching for e.g. "flint" will not find someone with the last name 'Flint', only flint artefacts. Once these entities have been found, they are indexed together with the full text in a search system. Detected time periods are automatically translated to year ranges and indexed as additional information. This makes it possible to search on year ranges instead of written timespans, tackling the problem of synonymy for this entity type. The system is called AGNES (Archaeological Grey-literature Named Entity Search), and it is available online at http://agnessearch.nl. In this project, 65,000 PDF documents were indexed, obtained from the DANS repository in 2017 (Brandsen et al. 2019).

To assess the usefulness and potential of AGNES, we perform a case study on Early Medieval cremation practices. In a recent article, Lippok (2020) has shown that Early Medieval cremations regularly occur next to the traditionally expected inhumation burials. The data for that article was compiled over three years through extensive literary study of published and unpublished books, PhD and master theses, and building on a career-long effort by Prof. Frans Theuws to map all Early Medieval sites in the Netherlands in the Rural Riches project (2021). This extensive study resulted in 77 sites containing cremation burials (Lippok 2020). Commercial reports were rarely consulted, due to the time-intensive process of surveying grey literature. Only after personal prompts by excavators indicating they had found Early Medieval cremations a few of these were added to the database.

Traditionally, cremation burials are presumed to disappear at the beginning of the Early Middle Ages (Effros 2003, 188; Fehr 2008, 77; Van Es, 1968, 8; James 1988, 141–142). Prior to 2020, their occurrence was never systematically investigated in the Netherlands due to their perceived incidental nature. When cremations were encountered in the past, they did not receive much attention as it was assumed they represented a burial rite that was on its way out of the burial repertoire (Van Es and Schoen 2008, 858). Cremations were mostly marginalised, the explanations for their occurrence was restricted to them representing individuals of different ethnic backgrounds, different religious persuasion or different social status from the majority of people that were inhumed (Lippok 2020).

The two burial forms were handled in exclusionary ways: where inhumation was prevalent, cremation was considered an anomaly. This dichotomous interpretation was countered by showing that inhumations and cremations occur most often together, even within the same grave context, rather than apart. There are further similarities: the use of the same material culture in cremations and inhumations, people of all ages and sex were both inhumed and cremated, and external funerary structures such as posts and mounds have been shown as similar too (Lippok 2020). Establishing the Early Medieval cremation ritual next to the inhumation ritual meant a shift in perspective, moving from homogeneous to heterogeneous burial repertoires.

Cremations are well suited as a case study for AGNES, as grey literature may contain information left out of traditional mediums such as books and papers. The results are relevant as they undermine an unbalanced focus on furnished inhumation graves: every single cremation grave contributes to the view that other ways of burial occurred and that archaeologists should not exclusively concentrate on what is considered the most prevalent type of burial.

This paper presents the latest version of AGNES, and describes a case study on Early Medieval cremations to show the usefulness of the system in archaeological research. The following research questions are addressed:

- How many Early Medieval cremations can we find in existing archaeological research reports that are currently unknown to a group of specialists in that area (the Rural Riches project), and how many known ones?
- What is the effectiveness of AGNES for retrieving relevant documents on this topic?
- To what extent does this new knowledge change our views on Early Medieval cremations?

Our contributions when compared to previous research are two-fold: (1) we present the first combined full text and entity search in the archaeology domain and (2) we provide the first systematic survey of Early Medieval cremations in Dutch grey literature. The creation of the search engine is work by Brandsen, and the research on cremations has been performed by Lippok.

2. Methods

In this section, we will first describe the search system, then explain the searching process, and end with a description of the evaluation.

2.1. AGNES

The latest version of AGNES was created based on a user requirement study (Brandsen et al. 2019) and an interface usability study (Brandsen et al. 2021a) with a small but representative group of Dutch archaeologists, to ensure the system is fit for purpose. An earlier version of AGNES, the initial prototype, is described in Brandsen et al. (2019). Since then, the interface was simplified and the accuracy of the NER has increased. Fig. 1 shows a screenshot of the current query interface, which allows users to search for any term (in the Query field), or search for particular archaeological entities (artefacts, contexts and species) and/or time periods, which are denoted by a start and end year. There is also a geographical search integrated (not pictured here). The system can be accessed at agnessearch.nl/search/agnesv2.

The entities were detected using machine learning models trained on expert-labelled data. Specifically, we have further pre-trained a BERT (Bidirectional Encoder Representations from Transformers) model, which is a deep neural language model. The BERT architecture is used for many Natural Language Processing (NLP) tasks, and it is currently achieving state of the art results on benchmark tasks in many languages (Devlin et al. 2019; Liu et al. 2019; Xiong et al. 2020). Once the generic

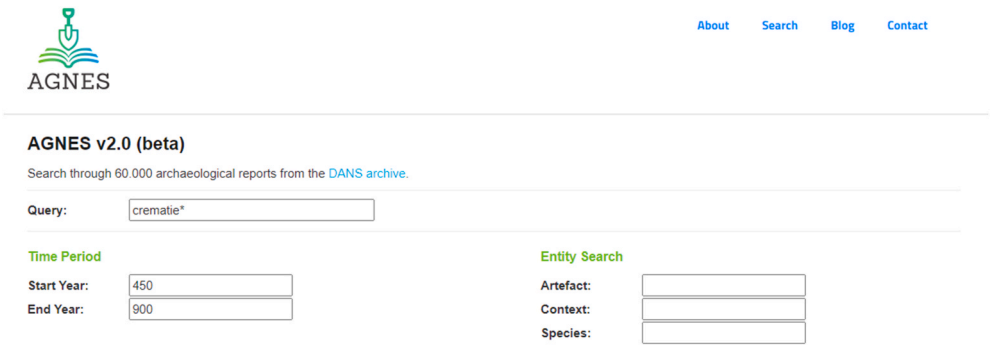

Fig. 1. Screenshot of the AGNES query interface (translated from Dutch).

language model is pre-trained, it can be fine-tuned for a specific NLP task, in our case NER.

A Dutch BERT model called BERTje has been released (De Vries et al., 2019). We further pre-trained the BERTje model on a collection of 65,000 Dutch archaeological documents obtained from DANS, training on four GPUs for about 24 h. The result is our ArcheoBERTje model. ArcheoBERTje outperforms the generic Dutch model BERTje, with an F1 score of 73.5% on our test set (Brandsen et al. 2021b).

ArcheoBERTje was used to detect entities in our entire document collection, leading to a total of over thirteen million entities. For time periods, the detected entities are translated to a year range (e.g. "Middle Ages" to 500–1500 CE), using a combination of regular expressions and dictionary lookups. All the entities, year ranges and full text of the documents are subsequently indexed in ElasticSearch, an open source search engine. The entities are stored as uncontrolled entities, i.e., they are not matched to thesaurus entries at the moment.

Currently, pages are used as the index unit, which means that if all the search terms occur on one page of a document, that document is returned as a relevant result. Indexing per page is not ideal, as some query terms might be split over multiple pages. It would be better to index per section, which is planned for future work.

An online user interface (see Fig. 1) can be used to query the free text, detected artefact, context and species entities, and the year ranges converted from time period entities. Results are displayed on screen, or can be exported to a CSV or GeoJSON file for further analysis.

2.2. Search process for our case study

To assess the usefulness of AGNES, a case study of Early Medieval cremations in the Netherlands was conducted, as introduced in section 2.1. To document the search process and results, the following process was used:

- The information need for this topic is defined based on prior knowledge
- A free search session using AGNES is conducted
- All entered queries are stored
- Results of the queries are stored in CSV format for further analysis, duplicate documents appearing in multiple result sets are removed

Once this was completed, we manually assessed the relevance of the retrieved documents and whether or not the result is already known from an earlier survey. This was done with the CSV export, which also contains links to page previews and the full document, to allow for detailed checking. This CSV approach for assessing the documents was preferred over using the search system's interface as there is a lot of overlap between the result sets of the different queries, thus reducing the number of documents that need to be assessed.

2.3. Evaluation: comparison to existing knowledge

To assess the usefulness of the results, the AGNES data was compared to an earlier survey of Early Medieval cremations in the Netherlands (Lippok 2020). That survey yielded 77 sites, based on comprehensive knowledge of the Early Middle Ages, but excluding grey literature. The reference database was compiled by going through all published, and some unpublished Early Medieval cemetery catalogi, comprehensive overview works of Early Medieval archaeology, master and PhD theses and other relevant material, such as site visits. The reference database is built on the career-long effort by Prof. Frans Theuws to make an overview of Early Medieval sites and has been in the making, specifically for cremations, for three years.

The comparison between the AGNES data and earlier survey data was made through assessing the relevance of the AGNES results, checking for the right time period, and if a cremation was actually found. After deselection of irrelevant records, the relevant AGNES records were cross checked with the list of known sites. A site is considered the same if they either have the same project name, or the same geographical coordinates.

3. Results

In this section, we describe the information needs and queries for the case study, show the results retrieved by AGNES, and compare these to existing knowledge.

3.1. Information needs and queries

An information need can be defined as a user's end goal in a specific search session (Hjørland 1997). For this case study, the information need is as follows: to find all mentions of Early Medieval cremations in grey

Table 1
All nine queries used to retrieve results, in the order in which they were issued. An English translation is given for Dutch terms. Asterisks (*) are wildcards.

Start Year	End Year	Free Text Query	English Translation	Number of documents retrieved
450	900	crematie	cremation	614
450	900	crematie*	cremation*	2335
450	900	verbrand menselijk bot	burnt human bone	24
450	900	brandstapel	pyre	73
450	900	brandstapel*	pyre*	84
450	900	urn	urn	508
450	900	knochenlager	bone bed	1
450	900	beendernest	bone bed	2
450	900	brandgrube	a pit containing pyre and cremation remains, covered by soil	8
			TOTAL	3035

literature, with the Early Middle ages being defined as the period 450–900 CE. In the search session, this resulted in the queries listed in Table 1. These queries have been thought of and constructed by the expert in this topic in a free search environment, without technical help, for a fair comparison to previous surveys. Synonyms for "cremation" are based on the expert's knowledge, no archaeological thesaurus was consulted.

The start and end year are entered in number fields (see Fig. 1), and are used to search through detected time period entities translated to absolute year ranges. The free text field was used to search for cremations, the entity search was not useful for these particular queries. The asterisks in the query column denote wildcards, meaning they match zero or more characters appended to the search term. So for example *crematie** will also match *crematieresten* (cremation remains), leading to more possible results. There are two German terms in the query column, *knochenlager* and *brandgrube*. This is because these German phrases are used interchangeably with their Dutch translations in reports.

3.2. Retrieved documents

When the results of these queries are combined and duplicate documents are removed, this leads to a total number of 2541 retrieved documents. The documents are ranked based on the free text query, the year range search is a boolean filter and as such does not influence the ranking For a full list of the results, please see the Zenodo repository[1] containing all data associated with this study (Brandsen and Lippok 2021). It took one person about 40 h to go through the list and mark the relevance for each document. Out of all the results, 54 documents are relevant to the information need, 41 documents are potentially relevant but unclear from the text, and 2446 documents are not relevant (see Table 2).

The large number of irrelevant results leads to a low precision of only 2.1%, with precision being defined as the fraction of relevant documents among the retrieved documents (Powers 2011). While the precision is low due to the large percentage of irrelevant results, this is not uncommon in systematic review studies. An example is the research by Bramer et al. (2013) in the biomedical domain, which achieved a precision of 1.9%, very similar to this study.

While having a higher precision would be useful, as it shortens the time needed to check the results, these kinds of tasks are recall-oriented: having as many relevant results as possible is more important than having a small number of irrelevant results. This has been documented for archaeologists specifically (Brandsen et al. 2019) and professional search more generally (Russell-Rose et al., 2018; Verberne et al., 2019). Unfortunately, the total number of relevant documents in the data set is not known, and therefore the recall can not be calculated. However, it is worth noting that out of the currently known 77 sites, AGNES has found 31, plus an additional 23 unknown sites.

When the irrelevant results are inspected (Table 3), it shows the vast majority are due to wrongly identified time periods (number 1 and 4), which can be attributed to NER errors. The other problems are mainly caused by specific types of sections in archaeological reports: lists of abbreviations and time periods, literature lists, etc (numbers 2, 3, 5, 8 and 9). Even though these pages will contain the correct search terms, they are always irrelevant as they do not describe an excavated cremation. A possible solution to this problem is described in section 4.3. Negations were expected to be a substantial problem, but with only 22 errors this does not seem to be the case for this information need.

Table 2
Overview of relevant, irrelevant and possibly relevant results. Relevant results are divided into previously known and unknown sites.

Type	Quantity
Relevant (Early Medieval cremation occurs in report) – known	31
Relevant (Early Medieval cremation occurs in report) – unknown	23
Possibly relevant (period or occurrence of cremation not explicit)	41
Not relevant (Early Medieval cremation does not occur in report)	2446
Total	2541

Table 3
Overview of the different categories of irrelevant results. Percentages are rounded to whole numbers.

Number	Type of error/type of irrelevant document	Quantity	Percentage
1	Wrong time period	1742	71%
2	Page listing abbreviations	235	10%
3	Page containing research plan (*Plan van Aanpak*[2])	198	8%
4	Unknown time period	122	5%
5	Page containing list of time periods	85	4%
6	Negation ("no cremation")	22	1%
7	Other	21	1%
8	Literature list	18	1%
9	Coring chart	3	0%

3.3. Comparison

The results of AGNES were compared with a database containing all Early Medieval cremations known from the earlier survey published by Lippok (2020). AGNES found 31 of the 77 known sites, and an additional 23 previously unknown sites containing Early Medieval cremations. 75 of the known sites were originally published in books and PhD theses, and two in excavation reports. However, the latter were published after 2017, which is after the data export from DANS, and as such can not occur in the AGNES results. In that sense it is surprising that AGNES found 31 of these sites published in books and PhD theses, as these are not included in the AGNES dataset. We assume these sites are mentioned in desk-based research reports that used the books as sources.

While 23 new sites might seem like a small number, this is a 30% increase over the existing knowledge. In the last 20 years, only 18 new Early Medieval cemeteries have been discovered in the Netherlands, and only nine sites containing cremation burials. With that as an indication of these site's scarcity, being able to add 23 sites that contain cremation burials is a major development.

The site of Hilvarenbeek is an excellent illustration of the added value of AGNES, it is a site completely missed by the earlier survey (Lippok 2020), yet it yielded C14 dated cremated remains dating between 550 and 620 CE (Claeys et al. 2012). Whilst this is an exciting find, it is unique in its novelty as it is the only cremation that was newly excavated and reported on in our list of 23. The other cremations are found in desk-research reports that scouted for archaeological sites surrounding the location of their study. Early Medieval cemeteries were most often mentioned there and are not often new finds. This means these sites are not newly excavated, and known to a few people, but practically impossible to find without AGNES.

Fig. 2 shows all the known and new sites, as well as all the sites found in the Dutch national cultural heritage database Archis for the search term "*Vroege Middeleeuwen*" (Early Middle Ages), to give an idea of the distribution of sites from this period. New sites containing cremations were located in areas known to yield cremations. Noord Brabant, Limburg, Overijssel, Zuid Holland, Gelderland, Drenthe, Friesland and Groningen have an additional one to six new sites. The sites of Castricum and Den Burg in North-Holland are located in a province that had up till now not yielded any Early Medieval cremations. Generally, Early Medieval sites are scarce in this province due to its geomorphological swampy nature in this period. A notable seven new sites containing cremations were added in the province of Drenthe, where previously

[1] Available at https://doi.org/10.5281/zenodo.3758085.
[2] A legal requirement in Dutch archaeology, the *Plan van Aanpak* describes the planned research methods.

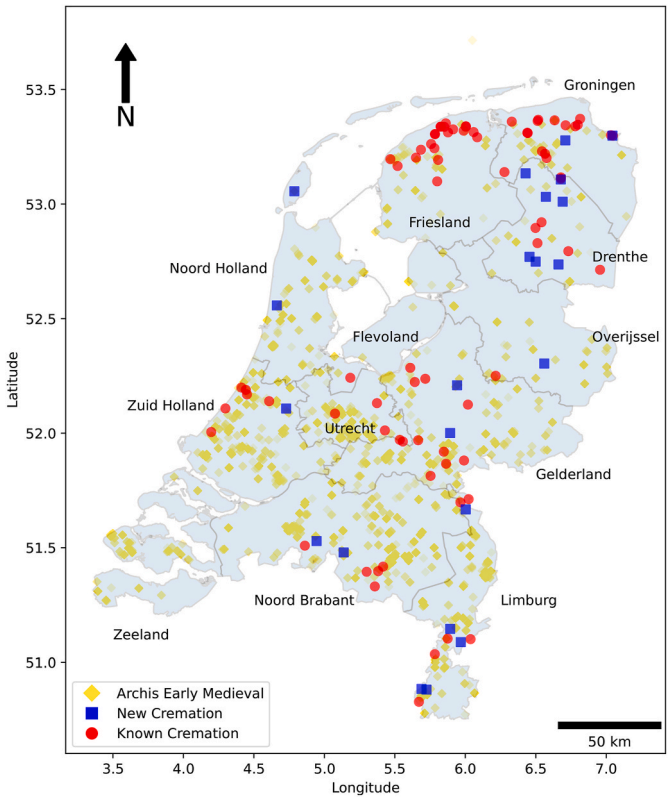

Fig. 2. Map of The Netherlands showing known sites (red circles) and previously unknown sites found with AGNES (blue squares). Yellow diamonds indicate known Early Medieval sites (with or without cremations) as recorded in the Archis system. Province names marked in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

only five were known.

Of the 23 new sites containing cremations, 16 are cemetery sites, one was found on a cemetery near a settlement and 6 were single cremations or urns, one of which probably indicates a larger cemetery. No additional context information about the other 5 single cremations was available.

4. Discussion

In this section, we discuss how the results from this case study affect our view on Early Medieval burial practices, the overall potential of AGNES, and suggest future work.

4.1. Archaeological significance

Finding an additional 23 Early Medieval sites containing cremation graves draws attention to this type of burial. Previously, cremations have been neglected in Early Medieval mortuary studies, their obscurity is emphasised, or they are altogether left out of discussions on mortuary rites (Lippok 2020). With renewed interest in cremations, they are not only shown to be more prevalent than before, additional information on the occurrence of cremation burials has come to light. Although the 23 new sites do not reflect this, previous research shows that Early Medieval cremations occur not only in cemeteries, but in settlements too, such as in Oegstgeest and Utrecht-Leidsche Rijn (Lippok, 2019, 2021). Because inhumations also occur in settlements, new questions concerning relations with the dead and their place in the landscape arise. Instead of being restricted to the cemeteries, away from the settlements, a more interactive relationship with the dead might be envisioned. The idea of interaction with the dead fits with the practice of re-opening graves, which consists of targeted retrieval of particular items and/or bones (Van Haperen, 2017). Cremations may have been kept above ground in containers and may have circulated for days or even years before being interred, as is suggested by Williams (2014) for Anglo-Saxon cremations.

In terms of geographical distribution, there are trends in the occurrence of cremations. The cremations in the north of the Netherlands, in the provinces Friesland, Groningen and Drenthe, conform to the prevalent known area of their occurrence. Both these and the cremations around the central Dutch rivers are often said to be restricted to these areas (Van Es and Schoen 2008, 858). This is accompanied by a heavily critiqued interpretative framework that emphasises ethnicity and religion as the main factor in the choice of burial form. Because of their seemingly incidental nature, cremations were thought to be of 'pagan' immigrants from the North, taking their 'traditional' burial practices with them (e.g. Hombert 1950, 100; Wamers 2015, 222). As for the more southern graves, they could fit in with the occurrence of cremation burials in and around the Scheldt area as described by Annaert (2018). Currently, Early Medieval cremations are also explained as belonging to immigrants in that area, with subsequent acculturation to account for cremations in later phases. However, through case studies of Lent and Elst in the Netherlands (Hendriks, 2013; Verwers and van Tent, 2015) Grobbendonk in Belgium (Janssens and Roosens 1963), and Lünen in Germany (Lehnemann 2008) it became apparent that the cremation and inhumation burials on these sites, and in a broader early medieval context, may be seen as more similar than was assumed before. Material culture in both types of graves are comparable and in the case of Grobbendonk even identical when looking at types of pottery and decoration (Lippok 2020). A case can be made for burial communities that had both burial methods at their disposal. The ethnic and religious explanations provided by previous research seem less likely, as they oppose the two types of burial in an interpretive sense whereas cremations and inhumations have been shown as comparable (Lippok 2020).

Finding 23 sites boosts the numbers of cremations, making it more compelling to argue that they should be included in considerations on the Early Medieval mortuary framework where they currently are neglected (cf. Lippok 2020). The occurrence of cremations actually helps us rethink Early Medieval burial practices. By incorporating the cremation rite in the Early Medieval mortuary framework, a more accurate portrayal of the archaeological reality is provided, and can therefore also give a more legitimate account of this time period.

4.2. Potential of AGNES for archaeological research

The results described in the previous section show that AGNES can be useful for synthesising archaeological research. It is not intended as a replacement for the currently available search tools, but as an additional method to gather data from hitherto underused sources. Of course AGNES is limited in that it can so far only search through reports deposited at DANS, and as such can not yet find information such as metal detectorist finds stored in the Archis system, or information from journals and books. What AGNES excels in though, is finding the by-catch of excavations often not mentioned in books and articles, and finding specialised or uncommon finds and contexts not mentioned in the metadata. This is illustrated by the fact that none of the new sites are returned when searching for "*crematie vroege middeleeuwen*" (cremation Early Middle Ages) in e.g. the DANS archive. These cremations are a good example of information that can relatively easily be found using AGNES, but would require manually searching through the entire collection of documents using the previously available systems.

While this case study has a very specific topic, AGNES can be used just as easily for other research questions. It is possible to combine any artefact, context, and species entity search with a free text search, a year range search and even geographical search (by drawing areas on a map). There are however a couple of limitations in the current system: it is not possible to search on multiple year ranges at once, and there is no

controlled vocabulary for the entity search. Both these issues will be tackled in a follow up project.

Besides being used for research questions in the Netherlands in Dutch documents, the system can also be adapted to other regions and languages (further described in the next section). The types of research questions are fairly similar across countries, most of them deal with What, Where and When questions (Jeffrey et al., 2009), and this would make it relatively easy to adapt to other regions as the entity types remain the same. The only prerequisites are labelled training data for the NER algorithm, and a region/language specific time period thesaurus to convert entities to year ranges.

Our work is similar to that of the STAR project (Tudhope et al., 2011), which shows the need for these kinds of tools in other regions. Although they focused more on more detailed metadata generation and AGNES focuses more on full text search combined with entity search, both projects used similar methods to find the What, Where and When aspects of grey literature. The STAR project also mapped entities to a thesaurus, leading to a controlled vocabulary of entities. This makes sense for their goal of interoperability, but in a free search scenario uncontrolled entities are more useful for users. However, we do plan to map entities to thesauri where possible in a follow-up project.

The results from this case study are promising, and help change long held views and biases resulting from an underrepresentation of cremations in published literature. We are optimistic that AGNES can help with other archaeological information needs as well, and will hopefully lead to a better understanding of the past.

4.3. Future work

However, in conducting this case study, some areas that could use improvement were found. The main issue is that while the relevant results are very useful, a relatively large amount of irrelevant results are also returned, making the checking process fairly time consuming. However, this is still much less time consuming than searching through the document collection manually, and similar amounts of irrelevant results are found in other systematic reviews. Around 70% of irrelevant results are due to the wrong time period being identified, an error propagation from the NER process. To solve this, more effort needs to be invested into making the NER process more accurate, specifically for the time period entities. Experiments with newer architectures such as RoBERTa (Liu et al., 2019) and LUKE (Yamada et al. 2020), as well as increasing the unlabelled training data, might increase the performance of the NER, and thereby decrease wrongly identified time periods. Our time period entity to year range conversion module will also need to be further tested and refined to further decrease false positives.

The second most common type of irrelevant document is due to specific sections that are not useful for searching, such as abbreviation, time period, and literature lists. A way to solve this would be to automatically detect these types of pages after the NER process, and either give terms on these pages a lower weight in the ranking algorithm, or to avoid indexing them altogether. This type of injection of domain or situational knowledge is already being successfully applied in automated detection in remotely sensed data (Verschoof-Van Der Vaart et al., 2020), and should improve results here too.

As mentioned previously, the current result evaluation process using the CSV export is time consuming, and as such it would be worth experimenting with ways that allow for interactive results checking. A solution would be a system where a query — or group of queries — can be saved in the online environment, which can then be further explored with links to page previews, and a method of marking results as relevant or not. After this process, the user can download a CSV export of just the relevant results. This should streamline the process.

Currently, the NER BERT model is only able to handle Dutch texts, but it is relatively easy to train a model for other languages. This would require annotated texts to train on, which can be produced to a sufficient quantity and quality in about 90 h of annotation (Brandsen et al. 2020). The actual training of the BERT language model would take about 24 h on 4 GPUs, and training the NER model another 4 h. In a follow up project, the system will be expanded to also handle English, German and possibly French texts, as well as diversifying the type of documents: including papers, books and theses, among others.

Besides these technical improvements, the archaeological side of this study also warrants further research. The 41 sites classified as 'possibly relevant' consist of sites where either the Early Medieval date is in question, or the occurrence of cremation is not explicit. The excavations at Park Leeuwenstein in Geldermalsen, for example, yielded cremations and inhumation dating from the Iron Age to the Carolingian period. More information is needed to assess if the cremations are Early Medieval. It is encouraging that verified Early Medieval cremations were found closeby, at Geldermalsen, Meteren de Plantage.This suggests that at least one of the 41 possible relevant sites contained another Early Medieval cremation. Given the small number of sites, it would be prudent to research all 41 sites to take away the doubt over their usefulness. Verification would involve a literature search on those sites and possibly contacting the authors of the report to ask for clarification.

5. Conclusions

In total, 23 additional sites containing Early Medieval cremations were found, when compared to a previous survey (Lippok 2020). This is a 30% increase on the total number of known sites before the study, and more than double the number of Early Medieval cemeteries discovered in the last 20 years.

The amount of information found is promising, but with a precision of 2.1% there are a large amount of irrelevant results that need to be manually assessed. However, this precision is similar to other systematic review studies, and the total amount of hours spent on assessing (~40 h) is much lower than it would be with other systems. More importantly, the time spent querying and assessing was deemed acceptable for the amount of information gained.

The additional 23 sites containing cremation graves further strengthens the importance of a heterogeneous perspective on Early Medieval burial repertoires. In the past, furnished inhumation graves were afforded most scholarly attention. The increasing number of sites containing cremation burials from the Early Middle Ages attests to a more heterogeneous burial repertoire. To understand Early Medieval communities, it is necessary to account for all of their burial practices. Understanding the occurrence of cremation practices will aid answering questions on heterogeneity, burial communities and change therein.

While some work needs to be done to further improve AGNES, the results presented in this case study are relevant and substantial, and the potential of the system seems promising for other information needs. We are confident that AGNES can become a useful tool to add to the archaeologists' searching toolbox, leading to more efficient and more detailed research.

Author statement

Alex Brandsen: Conceptualisation, Methodology, Software, Validation, Investigation, Data Curation, Writing - Original Draft, Visualization. **Femke Lippok**: Conceptualisation, Investigation, Resources, Writing - Original Draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The training of BERT models was performed using the compute

resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. The research on cremations is part of the Rural Riches project which is funded by the ERC under the ERC-Advanced Grant (Grant agreement No. 741340). We would like to thank the reviewers of this paper for their insightful and helpful comments.

References

Annaert, R., 2018. Het Vroegmiddeleeuwse Grafveld Van Broechem, Volume II Analyse. Habelt Verlag, Bonn.

Bramer, W.M., Giustini, D., Kramer, B.M., Anderson, P., 2013. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst. Rev. 2 (1), 115. https://doi.org/10.1186/2046-4053-2-115.

Brandsen, A., Lambers, K., Verberne, S., Wansleeben, M., 2019. User requirement solicitation for an information retrieval system Applied to Dutch grey literature in the archaeology domain. Journal of Computer Applications in Archaeology 2 (1), 21–30. https://doi.org/10.5334/jcaa.33.

Brandsen, A., Verberne, S., Wansleeben, M., Lambers, K., 2020. Creating a dataset for named entity recognition in the archaeology domain. Proceedings of The 12th Language Resources and Evaluation Conference 4573–4577. URL. https://www.aclweb.org/anthology/2020.lrec-1.562/.

Brandsen, A., Verberne, S., Lambers, K., Wansleeben, M., 2021a. Usability evaluation for online professional search in the Dutch archaeology domain. ArXiv preprint. ArXiv: 2103.04437.

Brandsen, A., Verberne, S., Lambers, K., Wansleeben, M., 2021b. Can BERT dig it? – named entity recognition for information retrieval in the archaeology domain. ArXiv preprint. ArXiv: 2106.07742.

Brandsen, A., Lippok, F., 2021. AGNES case study data. Zenodo Repository. https://doi.org/10.5281/zenodo.4737564.

Claeys, J., 2012. Onder de Gelderakkers een uitgestrekte meerperiodensite in Hilvarenbeek. Een inventariserend veldonderzoek door middel van proefsleuven (IVO-P). In: Baetsen, S., Drenth, E., Huizer, J., Jaspers, N., Kempkens, J., Lupak, T., Melkert, M.J.A. (Eds.), ADC Archeoprojecten Rapport 2613. https://doi.org/10.17026/dans-zvw-3mpv.

De Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., Nissim, M., 2019. BERTje: a Dutch BERT model. ArXiv preprint. ArXiv, 1912.09582.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies. Volume 1 (Long and Short Papers), 4171–4186. DOI: v1/N19-1423">10.18653/v1/N19-1423.

Effros, B., 2003. Merovingian Mortuary Archaeology and the Making of the Early Middle Ages. University of California Press, Berkeley.

Evans, T.N.L., 2015. A reassessment of archaeological grey literature: semantics and paradoxes. Internet Archaeol. 40 https://doi.org/10.11141/ia.40.6.

Fehr, H., 2008. Germanische Einwanderung oder kulturelle Neuorientierung? Zu den Anfängen des Reihengräberhorizontes. In: Brather, S. (Ed.), Zwischen Spätantike und Frühmittelalter. Walter de Gruyter, Berlin, pp. 67–102. https://doi.org/10.1515/9783110210729.2.67.

Habermehl, D., 2019. Over zaaien en oogsten, de kwaliteit en bruikbaarheid van archeologische rapporten voor synthetiserend onderzoek. Technical report. Amersfoort: RCE. URL. https://www.cultureelerfgoed.nl/publicaties/publicaties/2019/01/01/over-zaaien-en-oogsten.

Hendriks, J., 2013. Een merovingisch grafveld in het Lentseveld te Nijmegen-Noord. Evaluatie en selectierapport NLa14. Archeologie Gemeente Nijmegen (Archeologische berichten Nijmegen, briefrapport 1).

Hjørland, B., 1997. Information Seeking and Subject Representation: an Activity-Theoretical Approach to Information Science. Praeger, Santa Barbara, CA.

Hombert, P., 1950. Les sépultures mérovingiennes par incinération en Belgique. Rev. Archéol. 36, 96–102.

James, E., 1988. The Franks. Blackwells, Oxford.

Janssens, P., Roosens, H., 1963. Lijkverbranding en lijkbegraving op het merovingisch grafveld te Grobbendonck. Archaeologia Belgica 71, 265–272.

Jeffrey, S., Richards, J., Ciravegna, F., Waller, S., Chapman, S., Zhang, Z., 2009. The Archaeotools project: faceted classification and natural language processing in an archaeological context. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 367 (1897), 2507–2519. https://doi.org/10.1098/rsta.2009.0038.

Lehnemann, E.M., 2008. Das gräberfeld von Lünen-wethmar, Kr. Unna. Internationale archäologie 108. Rahden: Verlag Marie Leidorf GmbH.

Lippok, F.E., 2019. Een vroegmiddeleeuws graf aan de rand van de nederzetting. In: Norde, E.H.L.D. (Ed.), Nederzettingsresten uit de vroege middeleeuwen in het plangebied Leeuwesteyn Noord in Leidsche Rijn, Diemen. RAAP-rapport 3855, pp. 91–99.

Lippok, F., 2020. The pyre and the grave: early medieval cremation burials in The Netherlands, the German Rhineland and Belgium. World Archaeol. 52 (1), 147–162. https://doi.org/10.1080/00438243.2020.1769297.

Lippok, F.E., 2021. The early medieval graves of Oegstgeest. In: de Bruin, J., Bakels, C., Theuws, F. (Eds.), *Oegstgeest. A Riverrine Settlement in the Early Medieval World System*. Habelt Verlag, pp. 84–107. Bonn.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: a robustly optimized bert pretraining approach. ArXiv Preprint ArXiv 11692, 1907.

Powers, D., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology 2 (1), 37–63.

RCE, 2019. Archeologisch onderzoek - aantal onderzoeksmeldingen | Erfgoedmonitor. Rijksdienst voor het Cultureel Erfgoed, Amersfoort. https://erfgoedmonitor.nl/indicatoren/archeologisch-onderzoek-aantal-onderzoeksmeldingen. (Accessed 1 February 2021).

Richards, J., Tudhope, D., Vlachidis, A., 2015. Text mining in archaeology: extracting information from archaeological reports. In: Barcelo, J.A., Bogdanovic, I. (Eds.), Mathematics and Archaeology. CRC Press, pp. 240–254. https://doi.org/10.1201/b18530-15.

Rural Riches project, 2021. The rural Riches database. on 01-02-2021, URL. https://www.merovingianarchaeology.org/blog/about/the-rr-database/.

Russell-Rose, T., Chamberlain, J., Azzopardi, L., 2018. Information retrieval in the workplace: a comparison of professional search practices. Inf. Process. Manag. 54 (6), 1042–1057. https://doi.org/10.1016/j.ipm.2018.07.003.

Tudhope, D, May, K, Binding, C, Vlachidis, A, 2011. Connecting archaeological data and grey literature via semantic cross search. Internet Archaeology 30. https://doi.org/10.11141/ia.30.5.

Van Es, W.A., 1968. Grafritueel en Kerstening. Inaugural lecture Vrije Universiteit Amsterdam, Amsterdam.

Van Es, W.A., Schoen, R.P., 2008. Het vroegmiddeleeuwse grafveld van Zweeloo. Palaeohistoria 49/50 795–935. URL. https://ugp.rug.nl/Palaeohistoria/article/view/25161.

Van Haperen, M., 2017. Early Medieval Grave Reopenings in the Low Countries. Ph.D. thesis Leiden University, Leiden. https://doi.org/10.17026/dans-x6b-bvgj.

Verberne, S., He, J., Kruschwitz, U., Wiggers, G., Larsen, B., Russell-Rose, T., de Vries, A. P., 2019. First international workshop on professional search. ACM SIGIR Forum 52 (2), 153–162. https://doi.org/10.1145/3308774.3308799.

Verschoof-Van Der Vaart, W.B., Lambers, K., Kowalczyk, W., Bourgeois, Q.P.J., 2020. Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from The Netherlands. ISPRS Int. J. Geo-Inf. 9 (5), 293. https://doi.org/10.3390/ijgi9050293.

Verwers, W.J.H., van Tent, W.J., 2015. Merovingisch grafveld Elst-'t Woud. Rapportage archeologische monumentenzorg. Rijksdienst voor het Cultureel erfgoed, Amersfoort 223.

Vlachidis, A., Tudhope, D., Wansleeben, M., Azzopardi, J., Green, K., Xia, L., Wright, H., 2017. D16.4: final report on Natural Language processing. Technical report. URL. In: http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/D16.4_Final_Report_on_Natural_Language_Processing_Final.pdf.

Wamers, E., 2015. Francofurd 2. Das bi-rituelle Kinderdoppelgrab der späten Merowingerzeit unter der Frankfurter Bartholomäuskirche (>> Dom <<). Schnell and Steiner (*Schriften des Archäologischen Museums Frankfurt* 22(2)), Regensburg.

Williams, H., 2014. A well-urned rest: cremation and inhumation in early Anglo-Saxon England. In: Kuijt, I., Quinn, C.P., Cooney, G. (Eds.), Transformation by Fire: the Archaeology of Cremation in Cultural Context. University of Arizona Press, Tucson, pp. 93–118.

Xiong, Y., Huang, Y., Chen, Q., Wang, X., Ni, Y., Tang, B., 2020. A joint model for medical named entity recognition and normalization. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020), pp. 499–504. URL. http://ceur-ws.org/Vol-2664/cantemist_paper18.pdf.

Yamada, I., Asai, A., Shindo, H., Takeda, H., Matsumoto, Y., 2020. LUKE: deep contextualized entity Representations with entity-aware self-attention. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6442–6454. DOI: v1/2020.emnlp-main.523">10.18653/v1/2020.emnlp-main.523.