

Instructions for preparing NLP evaluation ‘gold standard’

Thank you for taking part in this exercise – your assistance with our research is greatly appreciated. Please try to complete the whole exercise but you are free to take as many small breaks as required in order to progress smoothly with the task. Your names and IDs will not be recorded with the data. Please ask the researchers if there is anything you do not understand, or if any of the colours pose difficulties.

1. Aim

Your aim is to highlight key archaeological concepts from a selection of OASIS grey literature document summaries. The purpose is to provide a set of reference data (the ‘gold standard’) against which the performance of an automatic Natural Language processing system will be evaluated. Therefore, it is important to be as systematic and as consistent as possible.

Different colours will be used to highlight four different types of concepts, using Microsoft Word, as explained below. We interested in highlighting individual concepts and also phrases that express meaningful connections between individual concepts. When highlighting a concept, please include any immediate modifiers in form of an adjective or adverb (*large, circular, narrow etc.*) Highlight lists of concepts as if a single multi-concept. It is not necessary to highlight any determiners (*a, the*) but if you do this will not affect the validity of the exercise. In addition to highlighting the individual concepts, please underline the immediate phrase that meaningfully connects the concepts.

2. The four types of archaeological concepts

- **Finds:** Physical objects that can be described as finds of archaeological interest. They may be man made (eg amphora) or naturally occurring (eg flint flakes). Finds can be described as items of a material nature that are units for documentation and have physical boundaries that clearly separate them from other objects. Examples of Finds are: *a coin, coins, finds, the bottle, a brick, the Aphrodite of Milos* as well as small finds, such as *bone fragments, flint flakes, etc.*
- **Time Periods:** All kinds of names or codes for historical periods. Time Periods may vary in their degree of precision and may be expressed relative to other time frames. Examples of Time Periods are *Early Roman, Medieval, 10th Century*, etc. Do not highlight exact dates such as 1749, 6 of August and 10/09/1980 etc.
- **Contexts:** Spatial elements that constitute an individual archaeological unit of excavation. For this exercise, include both primitive contexts and larger groupings of contexts. Examples of Contexts are *pits, a cut, the deposit, context* (itself) as well as larger contexts, such as *enclosures, post-holes, hearths, a well, the floor, a structure, buildings, roads, etc.*
- **Materials:** Materials that have an *archaeological* interest and are associated with physical objects (finds). Examples of Materials are *iron, copper, charcoal etc.* A word should be treated as a Material if it modifies a Find but should be treated as a Find otherwise. For example, *the brick* is a Find, whereas *the brick oven* is a Material followed by a Find. Similar examples are *flint flakes* and the *pottery fragments* versus simply *flints* and *pottery*.

3. Further guidelines

Please highlight with the following principals in mind.

1. **Negation Detection:** Concepts or phrases that are negated should NOT be highlighted. For example, *No context was found to contain pottery* should NOT be highlighted. Since *context* is not highlighted because of negation, this means that *pottery* is also not highlighted
2. **Relevance:** Consider how relevant the concept is to the overall discourse. Do not highlight concepts that are obviously irrelevant to the archaeological focus of the OASIS report or carry irrelevant word senses. For example the word *Cut* which can be an archaeological or a verb, the later word sense should not be highlighted.
3. **Disambiguate** between Finds and Materials. For example the term 'brick' can either refer to a material i.e. (a brick wall) or to a physical object i.e. (a brick found in a context). You should decide on the conceptual alignment of terms that can be either materials or physical objects.
4. **Endings:** You should also highlight plural, gerund other endings when applicable. For example highlight 'bricks' 'panning', etc.
5. **Phrases:** You should consider conjuncted and adjectival phrases. For example you should consider conjunctions of the kind 'Early Roman to Late Roman', 'Pottery and brick', as well as 'worked flint', 'small finds' etc.
6. **Underline meaningful connections between concepts** by marking boundaries of phrases in text that normally begin with a single concept and end with another concept. It does not matter which concept begins and which ends the phrase. For example, highlight both phrases; *Roman period Find* and *find of Roman period*. Also highlight, phrases with more than two concepts. For example, underline *Iron nails of the Roman period* involving a Material (*Iron*), a Find (*nails*) and a Time Period (*Roman period*). For example, underline *Deposit containing iron finds of the Roman period* which also involves an archaeological context (*Deposit*).

Underline as many concepts as are meaningful in a phrase. The connecting phrases can be very short with just two concepts next to each other (eg Roman Find) or the connections can be larger, spanning to a dozen or more words. For example you should underline as one phrase '*Iron finds of the Roman period have been associated with the pit*' ,

4. Colour Coding

- Find {Grey 50% : #999999}
- Time Period {Turquoise : #00FFFF }
- Archaeological Context {Bright Green: #00FF00 }
- Material {Red : #FF0000 }
- Underline connection between concepts

*Note that if problem distinguishing colours talk to the researcher and choose a different colour.

5. Meaningful connections between concepts

1. Time Period **and** Find
i.e. **Mediaeval Pottery**
2. Time Period **and** Archaeological Context
i.e. **Roman deposits**
3. Archaeological Context **and** Find
i.e. **Ditch containing coins**
4. Material **and** Find
i.e. **Copper alloy artefacts**

More complex Examples

1. **Two pits, a posthole and a linear cut**, which are broadly dated from the **Neolithic period to the Late Bronze Age**
2. A quantity of **human bone** was recovered from its **fill**
3. **Ceramic** artefacts included **pottery** sherds, **roof tiles** and **bricks** all dated to the **Roman period**
4. **deposit** was medium brownish grey **silty sand** that also contained frequent **charcoal**
5.a broad range but three fragments off **glass bottles** in (13/007) were dated to **the late 18th early 19th century** and this is probably a.....
6.spits 35 **Postmedieval and possibly medieval deposits**, to be taken down carefully in one to three spits, down to c.11.1010.75m OD
7.a total of 7 **copper alloy artefacts** were recovered from various **Bronze Age contexts** during the excavation/evaluation

***NOTE** Example 5 mentions archaeological finds by a reference number for example (13/007). Reference numbers for finds and contexts should **not** be highlighted.