

NLP-Werkstatt

SS2017

XML

Hausaufgabe – Abgabe bis 24.05.2016, 14:00 Uhr

Aufgabe 1 Gegeben sei der Wörterbucheintrag *Interesse* vom DWDS (Das Digitale Wörterbuch der deutschen Sprache) <http://www.dwds.de/?view=1&qu=Interesse>¹².

The screenshot shows the DWDS entry for 'Interesse'. It includes the word 'Interesse' with its grammatical information (neutr., -s, -n) and its Latin origin. The entry is divided into four numbered sections: 1. 'aufmerksame geistige Teilnahme ohne Plural' with examples like 'großes, geringes, wenig, kein, lebhaftes, leidenschaftliches, reges, wachsendes Interesse an etw., jmdm., für etw., jmdn. haben, zeigen'; 2. 'Nachfrage, Neigung zum Kauf ohne Plural' with the example 'für diese Waren besteht großes, nur geringes Interesse'; 3. 'a Belange, Absichten, Bestrebungen nur im Plural' and 'b im Interesse einer Sache, Person' with various examples; 4. 'veraltet Zinsen, Gewinn nur im Plural' with the example 'jetzt muß er mir für das Geld ... noch Interessen zahlen'. A 'Dazu:' section lists related terms like 'Arbeiterinteresse', 'Berufsinteresse', etc. The bottom of the window shows the version (0.4.23) and source (WDG | Artikeltyp: Vollartikel).

Bauen Sie ein Wörterbuch mit *Interesse* als einen der Einträge, wobei Sie folgendes Gerüst erweitern sollten:

```
1 <?xml version='1.0' encoding='utf-8'?>
2 <dictionary language="German" source="http://www.dwds.de">
3 ...
4 </dictionary>
```

Bilden Sie für den Eintrag (mindestens) folgende Informationen in einer XML-Struktur ab:

- Lemma
- morphologische Eigenschaften
- Lesarten - für jede Lesart und Unterlesart:
 - Definition (Paraphrasierung) der (Unter-)Lesart
 - Beispielverwendung (Kollokationen mit dem aktuellen Eintragslemma)
- + geben Sie dem Eintrag und jeder (Unter-)Lesart eine ID

¹Abbildung anhand einer früheren Darstellung der Wörterbucheinträge.

²Der Originaleintrag enthält auch noch weitere Informationen, die man sieht, wenn man auf die "+"-Zeichen klickt und dadurch zusätzliche Informationen einblendet. Diese muss man jedoch für diese Aufgabe nicht beachten.