

1000 danske talemåder og faste udtryk

Delaflevering 1

Det Danske Sprog- og Litteraturselskab har udviklet et datasæt som indeholder samlet 1000 danske talemåder og faste udtryk med overførte betydninger (herefter omtalt som talemåder)¹. Datasættet er baseret på Den Danske Ordbogs samling af faste udtryk med tilhørende definitioner.

Mere om datasættet

Filen *talemaader_leverance_1.csv* indeholder tre tab-separeret kolonner:

| | |
|------------------------|--|
| udtryk_id | Id-nummer for talemåden. Vha. dette Id-nummer kan datasættet linkes til udtrykket i Den Danske Ordbog |
| talemåde_udtryk | Selve talemåden. Ved varierende former (fx <i>der er forår (øretæver, amoriner, ..) i luften</i>) har vi valgt én form |
| ddo_definition | Definition fra Den Danske Ordbog. Ved varierende former har vi tilpasset definitionen så den afspejler den valgte form. |

Datasættet udgives med en CC-by licens, hvilket betyder at Det Danske Sprog og Litteraturselskab skal krediteres ved brug.

Kontaktperson til datasættet

Nathalie Hau Sørensen [nats \(snabel-a\) dsl.dk](mailto:nats (snabel-a) dsl.dk)

Delafleveringen er første aflevering ud af to til Digitaliseringsstyrelsen. Delaflevering to indeholder tre falske definitioner til hver af de 1000 talemåder.

¹ Nogle udtryk er markeret 'talemåde' i Den Danske Ordbog, men ikke alle hvor det kunne være relevant. Datasættet indeholder derfor ikke kun disse, men inkluderer også faste udtryk med overførte (metaforiske) betydninger, særligt udtryk der indeholder et centralt ord i dansk. Fælles for alle udtrykkene er at man ikke kan udlede betydningen ud fra enkeltordenes betydning.