

Book review corpus from Bokelskere.no

This corpus is a dump of user generated book reviews and discussions from Bokelskere.no ("book lovers"), a web community where users review and discuss new and old literature, both fiction and non-fiction.

The corpus is structured as a JSON Array where each object corresponds to a review or comment to a review on Bokelskere.no. Each post/object has the following fields:

- "post_id": unique identifier for review
- "date": date when the review was posted
- "user_id": unique identifier for the user
- "isbn13": ISBN for the the reviewed book
- "post_title": title of review
- "text": review
- "score": evaluation (from 1-6, where 6 is the best)
- "main_title": title of reviewed book
- "author": author of reviewed book
- "parent_id": identifier of review which has been commented upon

The corpus contains approximately 219.000 posts/objects, and 1.5 million word tokens (in the «text»-field).

Bekomtalekorpus frå Bokelskere.no

Dette korpuset inneheld ein dump av brukargenererte bokomtalar og diskusjonar frå Bokelskere.no, ein nettstad der brukarane skriv omtalar av og diskuterer nye og eldre bøker, både skjønnlitteratur og fagprosa.

Korpuset er på JSON-format, der kvart objekt svarar til ein omtale eller ein kommentar til ein omtale på Bokelskere.no. Kvart objekt inneheld dei følgjande felta:

- "post_id": unik identifikator for omtalen
- "date": dato når omtalen blei posta
- "user_id": unik identifikator for brukaren
- "isbn13": ISBN-nummer for den aktuelle boka
- "post_title": tittel på omtalen
- "text": omtalen
- "score": evaluering (terningkast 1-6)
- "main_title": tittel på boka
- "author": forfattar av boka
- "parent_id": identifikator for ein omtale som er kommentert

Korpuset inneheld omlag 219.000 postar/objekt, og 1,5 millionar ord (i «text»-feltet).</description>

Bokanmeldeskorpuser fra Bokelskere.no

Dette korpuser inneholder en dump av brukergenererte bokanmeldelser og diskusjoner fra Bokelskere.no, et nettsted der brukerne skriver omtaler av og diskuterer nye og eldre bøker, både skjønnlitteratur og fagprosa.

Korpuser er på JSON-format, der hvert objekt tilsvarer en omtale eller en kommentar til en omtale på Bokelskere.no. Hvert objekt inneholder følgende felter:

- "post_id": unik identifikator for omtalen
- "date": dato når omtalen ble postet
- "user_id": unik identifikator for brukeren
- "isbn13": ISBN-nummer for den aktuelle boken
- "post_title": tittel på omtalen
- "text": omtalen
- "score": evaluering (terningkast 1-6)
- "main_title": bokens tittel
- "author": bokens forfatter
- "parent_id": identifikator for en omtale som er kommentert

Korpuser inneholder omlag 219.000 poster/objekter, og 1,5 millioner ord (i «text»-feltet).