WATCH · SANSKRIT COMPUTING / STUDY NOTES

Machine Translation and LLM-Powered Grammatical Explanation for Sanskrit - 7th ISCLS, Auroville

Study notes by Sanskrit AI Universe, based on this video

Video by Sebastian Nehrdich and colleagues (7th ISCLS) · Sanskrit Research Institute · 2 Mar 2024

Not the creator's own slides

About these notes

— Official title: Machine Translation and LLM-Powered Grammatical Explanation for Sanskrit - 7th ISCLS, Auroville

— Speaker and channel: Sebastian Nehrdich and colleagues (7th ISCLS) · Sanskrit Research Institute

— Published 2 Mar 2024 · Running time 18 min 32 s · https://www.youtube.com/watch?v=N1Kth8zIkW0

— Official description: Machine Translation and LLM-Powered Grammatical Explanation for Sanskrit Sebastian Nehrdich, David Allport, Jivnesh Sandhan, Kurt Keutzer

— Catalogue caution: Results as of early 2024; models have moved on.

— These pages are study notes drawn from the video's own title, description and captions, plus the journal's already published Watch note. They are not a transcript and not a substitute for watching the film.

— The journal treats videos as leads, not as sources you should cite on their own. Check the papers and editions behind a claim.

— Caption source: English (auto-generated). Auto-captions can mis-hear words; obvious proper-name spellings were given in IAST where the journal already uses them. Distinctive phrases follow the captions closely.

From the talk · 0:00–3:38

— For classical s can you so I will first can you keep it keep your presentation on the presentation mode or full screen yeah oh yeah sure you mean like.

— The architecture of the grammatical pre-processing step that we are proposing the machine translation and I will also give one example of the system in action and.

— But just as an overview there's Kurt executive director then I am the technical director David alport is working on Sanskrit data collection and then we have j n.

— That is a technique that can improve the quality of the translation here and then with all these different information that we have gathered.

From the talk · 3:40–7:29

— And they are prone to show hallucinations imprecise renderings of grammar and so on and I think this is familiar to many people in the audience.

— In an easy to understand manner few short examples actually can further increase the quality of translation when high quality data is available so.

— We combine all of these different aspects grammatically processing pre-generated machine translation and few short examples we can carry a large language model with a strong reasoning capabilities.

— In order to generate high quality Sanskrit translations together with a easy to understand explanation of the grammatical structure of the Sanskrit and.

From the talk · 7:31–11:13

— This conl format needs to be converted into some kind of natural language so that the model can actually work with that coming to the draft machine translation steps so.

— For epic and Pana text is pretty good for scriptures it's not as good for trees we are doing not so bad but it could be better poetry.

— I think it's still a lot of room of improvement Bic material of course as well the coverage is not very far at the moment so but all in all.

— We have managed to gather about 22% of the available data and our reference Corpus here is the G Corpus of University of betting so of that cpose.

From the talk · 11:16–14:58

— I can yeah okay it's okay continue please okay regarding further work so the Outlook is of course collecting more parallel sent data is very important because the better.

— The reference data is the better the training data for the draft translation system is the more reliable this whole pipeline is going to work so.

— When I saw your were workflow I noticed that translations are coming from different sources and the m ological information is coming from a different source.

— So yeah do you U have you don't you think there will be cases where the translation is different and morphological information that you are displaying is totally irrelevant.

From the talk · 15:00–18:29

— Because those conflicts in grammatical annotation and translation from the data side of you are very difficult to solve yeah because that would require basically retranslating the.

— The Sanskrit based on our grammatical passing or it would require to reannotate the Sanskrit based on our English translations and both is probably something that is well beyond what.

— That will undoubtedly emerge later on yeah so here one suggestion is to plug in the diction since you have the root words and you have the dictionary and then use.

— So just to quickly respond to the dictionary so we have tried that and the problem at the moment is that it's actually giving too much information to the pump of.

Sources and credit

— Study notes by Sanskrit AI Universe, based on the video “Machine Translation and LLM-Powered Grammatical Explanation for Sanskrit - 7th ISCLS, Auroville”. These are not the creator's own slides.

— Video by Sebastian Nehrdich and colleagues (7th ISCLS), posted on the YouTube channel Sanskrit Research Institute.

— Watch the video: https://www.youtube.com/watch?v=N1Kth8zIkW0

— Captions used: English (auto-generated) (en), retrieved on 24 November 2026 via a public Invidious instance. Auto-captions can mis-hear words; obvious proper-name spellings were given in IAST where the journal already uses them.

— Also used: the official video title and description, and the Watch catalogue note already published in this journal (presenter and why the video is listed).

— If the Founding Editor later supplies his own deck for this video, replace the PPTX and PDF in src_assets/slides/ (see the journal README, “How to swap in RK's own slides”).