WATCH · SANSKRIT COMPUTING / STUDY NOTES

Sanskrit & Computational Linguistics — Dr Pawan Goyal, IIT Kharagpur

Study notes by Sanskrit AI Universe, based on this video

Video by Pawan Goyal (IIT Kharagpur) · Institute for Science and Spirituality Delhi · 28 Sep 2021

Not the creator's own slides

Sanskrit AI Universe

About these notes

— Official title: Sanskrit & Computational Linguistics | Dr Pawan Goyal, IIT Kharagpur

— Speaker and channel: Pawan Goyal (IIT Kharagpur) · Institute for Science and Spirituality Delhi

— Published 28 Sep 2021 · Running time 1 h 02 min · https://www.youtube.com/watch?v=m6IBfoE0hzA

— Official description: The design and implementation of computer-aided processing tools is of paramount importance to analyse the enormous store of knowledge and literature available as Sanskrit text. The talk will describe our ongoing efforts in developing a digital infrastructure for processing Sanskrit by effectively combining state-of-the-art machine learning techniques and traditional linguistic knowledge from Sanskrit. #scienceandspirituality #sanskrit #issdelhi #vedicscience About the Speaker: Dr. Pawan Goyal j

— 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 official description

— The design and implementation of computer-aided processing tools is of paramount importance to analyse the enormous store of knowledge and literature available as Sanskrit text.

— Prior to that, he was working at INRIA Paris-Rocquencourt as a post doctoral fellow with Prof. Gérard Huet on The Sanskrit Heritage Site.

— D. from Intelligent Systems Research Centre, Faculty of Computing and Engineering, University of Ulster, UK.

— The topic of his PhD dissertation was "Analytic Knowledge Discovery Techniques for Ad-Hoc information Retrieval and Text Summarization".

— His main research interests include Text Mining, Natural Language Processing, Information Retrieval and Sanskrit Computational Linguistics.

From the talk · 0:10–6:40

— Yes, such So, we have So, we have our own ideas about yes, why Sanskrit is important.

— But Sanskrit is the only language that has a very formal grammar that's that was written by uh Maharishi Pāṇini in 2,500 years ago.

— And this gives a complete description of the language like how do you generate different words in the language?

— And that has given so many theories that I because I work in the field of NLP, natural language processing, I can easily relate.

From the talk · 6:43–12:51

- This is composed This is composed of two different words and they just come together. come together.
- So as we know most of the Sanskrit literature, most of the things that we read, they're in poetic format, right?
- Uh but when in schools and colleges this is taught, they also tell what is its corresponding prose format.
- So like in English when I say I say, "Okay, I uh I say I go to market." Like I follow a very specific ordering.

From the talk · 12:53–18:53

— So, can we use some natural language processing methods or machine learning methods to automatically create these readers from the text?

— People in the world have been doing things for other languages, can we not use those things directly for Sanskrit?

— And then, other thing that's that drives machine learning is that you need to have a lot of data that is annotated, right?

— So, you need a lot of labels and you can train your system and this will give you an output.

From the talk · 18:55–25:13

- So, if you try the same technique, you will not be able to find the word the phoneme u, right? right?
- So, this is an actual word actual words in poetry in poetry and this is in the prose format, right?
- Yes, there is a subject, yes, then the uh This is the word, yes, subject and word, and this is the adjective.
- So, then I need to make some changes to the models that are available for the languages.

From the talk · 25:16–31:25

— So that is where I know only when I say this is a Bahuvrihi compound that refers to someone else, not from the two components.

— So with the So with the help of the students and So we came up with this graph-based framework for structure prediction task in Sanskrit.

— So you give us give a particular input and this will give all the possible things in a very summarized format.

— Now uh So that means annotators can come in and they can manually select what is the correct solution.

From the talk · 31:27–37:37

— I can have another problem that, "Okay, in this text in this text, uh what are the individual What are the names, right?

— We said, "Okay, I'll take it as a structured prediction framework where my input is structure, output will also be a structure, okay?

— Like, "Okay, uh there's a lot There's a lot of data available that is not tagged, but the data is available.

— Ta plus na also give you nana na and ma plus na also give you So, they are possibilities.

From the talk · 37:39–43:23

— So in graph theory maximal clique is This is a subgraph, this is a connected subgraph that is of maximal size.

— In machine learning for training, what we say, "Okay, this is a correct thing, this is an incorrect thing." Right?

— We say, "The correct thing should have the minimum energy, incorrect thing should have the higher energy." Because energy should be minimized.

— If I know that, "Okay, that, "Okay, this both of these are kartas, that cannot happen in the same sentence.

From the talk · 43:25–49:16

— And this model, once it is learned by some training data, it gives me a final So, that is the gist, right?

— So there he has a sentence and for each sentence what is the correct morphology correct segmentation that is available.

— For each word what is the output and word plus we say okay, the word and also the morphology both are correct.

— So if you have to publish we have to show that yeah this works better than all the possible available things.

From the talk · 49:18–56:06

— So with that I will thank all the participants and also the organizer for giving me this opportunity and yeah open to any questions.

— So with the help of this kind of things uh we can have a glimpse of uh what is there in our tradition and sundry details.

— So I have a question that uh Sanskrit the Sanskrit uh the order of words it doesn't really matter.

— So I think uh your works can also be would be kind of a superlative or superset of uh approaches to decode other languages as well.

From the talk · 56:09–1:02:29

— what is the future prospect in relation in relation with the Sanskrit and the computer language.

— So, it's not of so much importance to scholars because they already can do on their own.

— We are If you are If you are native speakers, we know oh what is the actual meaning of this word in this context.

— Yeah, I got it and I just want to tell you that I learned the language the language uh Sanskrit from your institute from the Anuradha Choudhary ji.

Sources and credit

— Study notes by Sanskrit AI Universe, based on the video “Sanskrit & Computational Linguistics | Dr Pawan Goyal, IIT Kharagpur”. These are not the creator's own slides.

— Video by Pawan Goyal (IIT Kharagpur), posted on the YouTube channel Institute for Science and Spirituality Delhi.

— Watch the video: https://www.youtube.com/watch?v=m6IBfoE0hzA

— 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”).