Vedavaapi
वेदवापी

The Nourishing Stepwell of Indic Knowledge

Online Portal for modern seekers to easily
Assimilate, Exchange and Apply
India's Native Scientific Knowledge

Vedavaapi Overview

- Charter: *Remove Barriers preventing Modern Seekers from understanding and applying Indic Knowledge, using Information Technology*

- Vedavaapi Foundation (Registered Non-profit Public Trust in India):
  - Key tenets: Not-for-profit, free access to IP for non-commercial and public use
  - Income sources: Donations (cash and infrastructure), support contracts

- Products and Services
  - Vedavaapi.org: Web service to access, share and process Indic Knowledge content
  - Services: Tools for search, analysis and annotation of digitized Indic texts, media

- Vedavaapi Labs (Tech centers):
  - Can be standalone or collocated with an educational institute

Indic Knowledge: Why Now?

- Individual growth is highly skewed under modern education
  - High intellectual focus, low emotional balance, internal & external unrest
  - Its effects on society are dire
- World needs a holistic alternative, now more than ever
- Indic wisdom has the answers
  - Time-tested, ever-relevant, thorough

Indic Knowledge: Areas of Unique Value

Subjective Knowledge:
Integral psychology, consciousness studies

- Where useful: Education, Social organization, Leadership & Administration, Wellness and peace
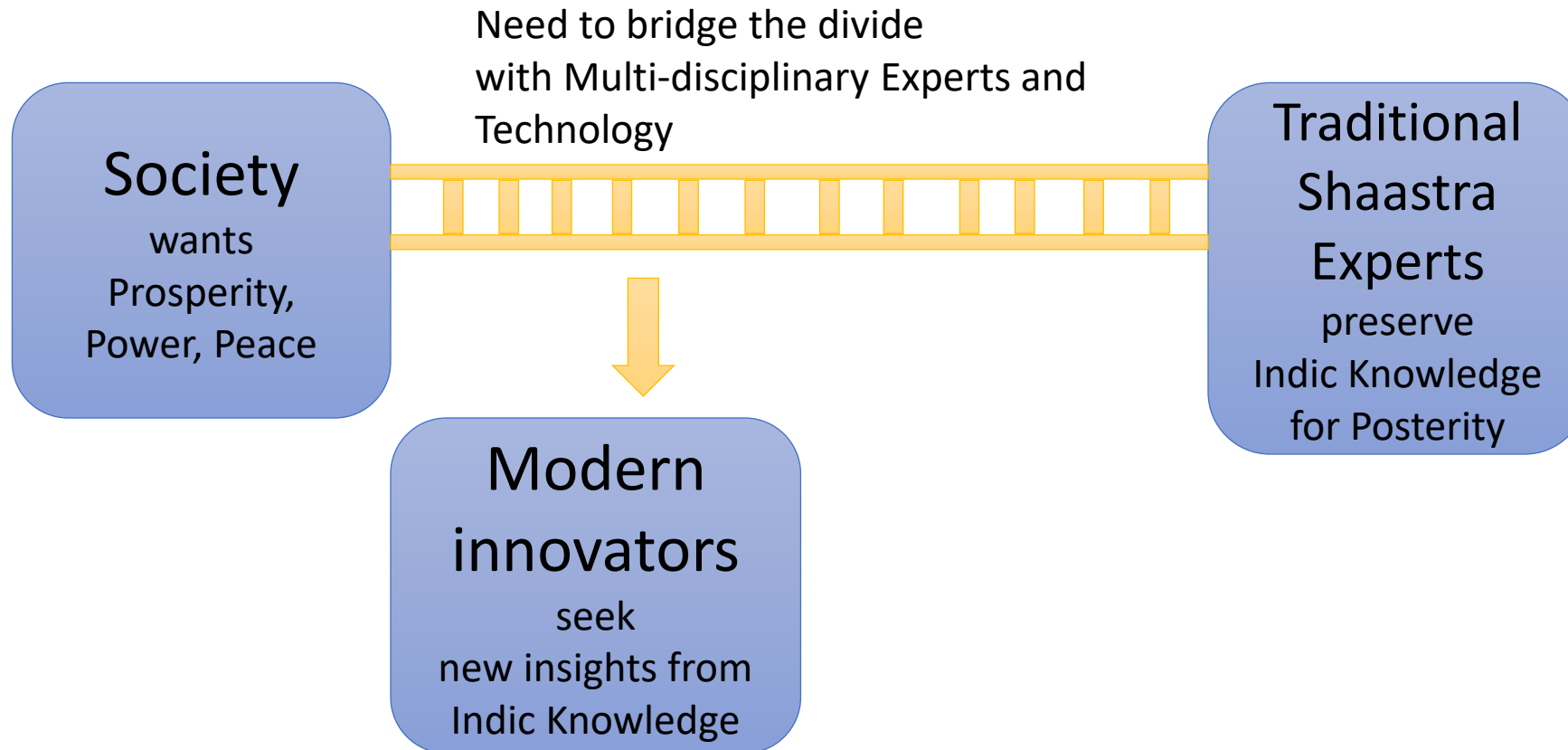
Objective Knowledge:
of Nature, its resources, properties, utilization

- Where useful: Physical sciences, engineering, technology

Knowledge of Knowledge:
codifying, articulating, processing

- Where useful: Linguistics, Math, logic, Knowledge mining, Automation Science
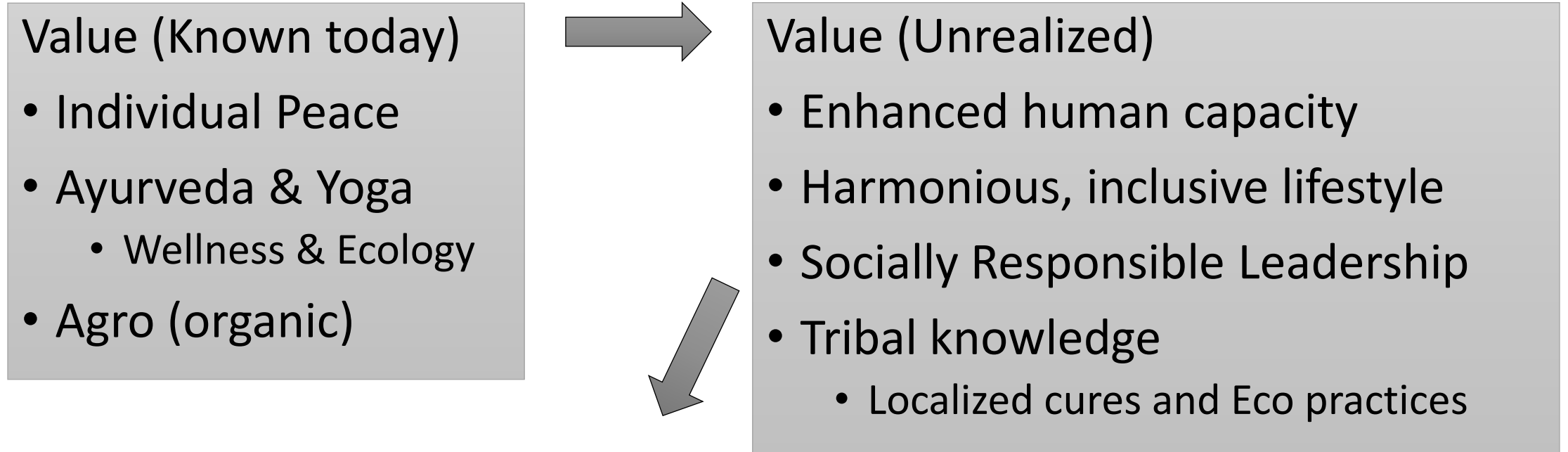
Indic Knowledge: Current State

Need to bridge the divide
with Multi-disciplinary Experts and
Technology

Society
wants
Prosperity,
Power, Peace

Traditional Shaastra Experts
preserve
Indic Knowledge
for Posterity

Modern innovators
seek
new insights from
Indic Knowledge

Indic Knowledge: Inhibitors to Adoption

- Lack of Employability <- Unclear economic value

- Lack of Pride in Indic Heritage <- Ignorance, degradation, foreign rule

- Lack of Samskrit literacy <- Unnatural teaching via translation

- Lack of Modernized Access <- Method unsuited to modern minds

- Lack of Familiarity with Native Indic thought culture
  - Because India's education got severed from its roots for generations

Indic Sciences: Unlocking New Value

Value (Known today)
- Individual Peace
- Ayurveda & Yoga
  - Wellness & Ecology
- Agro (organic)

Value (Unrealized)
- Enhanced human capacity
- Harmonious, inclusive lifestyle
- Socially Responsible Leadership
- Tribal knowledge
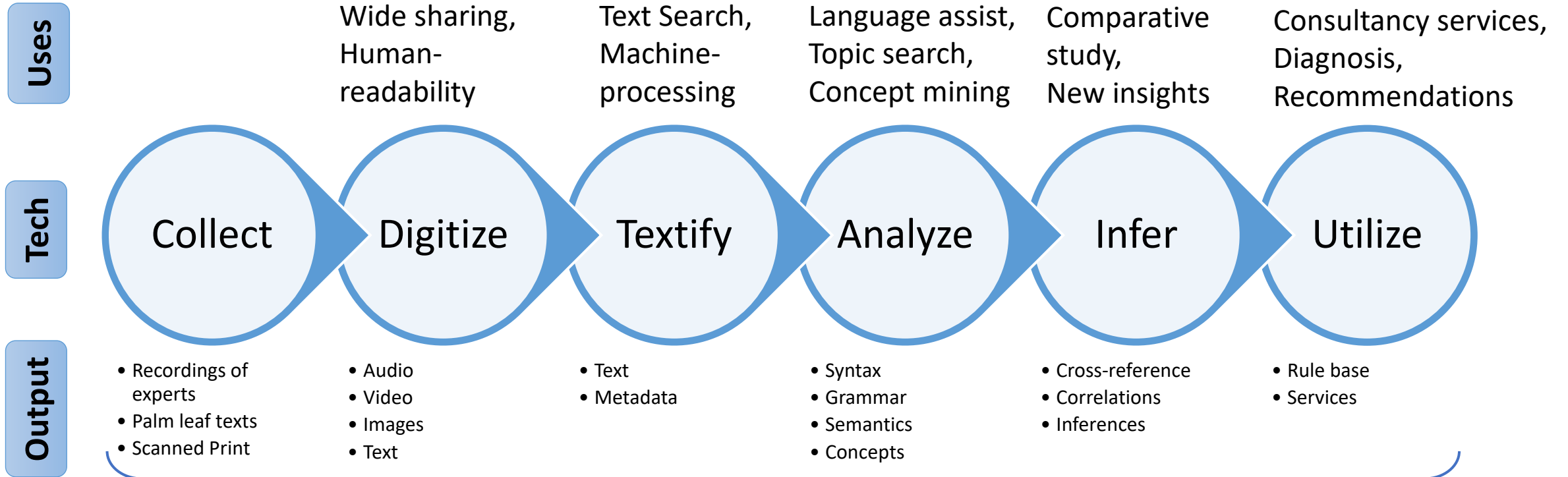 - Localized cures and Eco practices

How to Realize?

- Familiarize modern students with Native Indic Thought
- Launch a platform to easily imbibe and exchange Indic Knowledge
- Spur Open-minded, Comparative Research
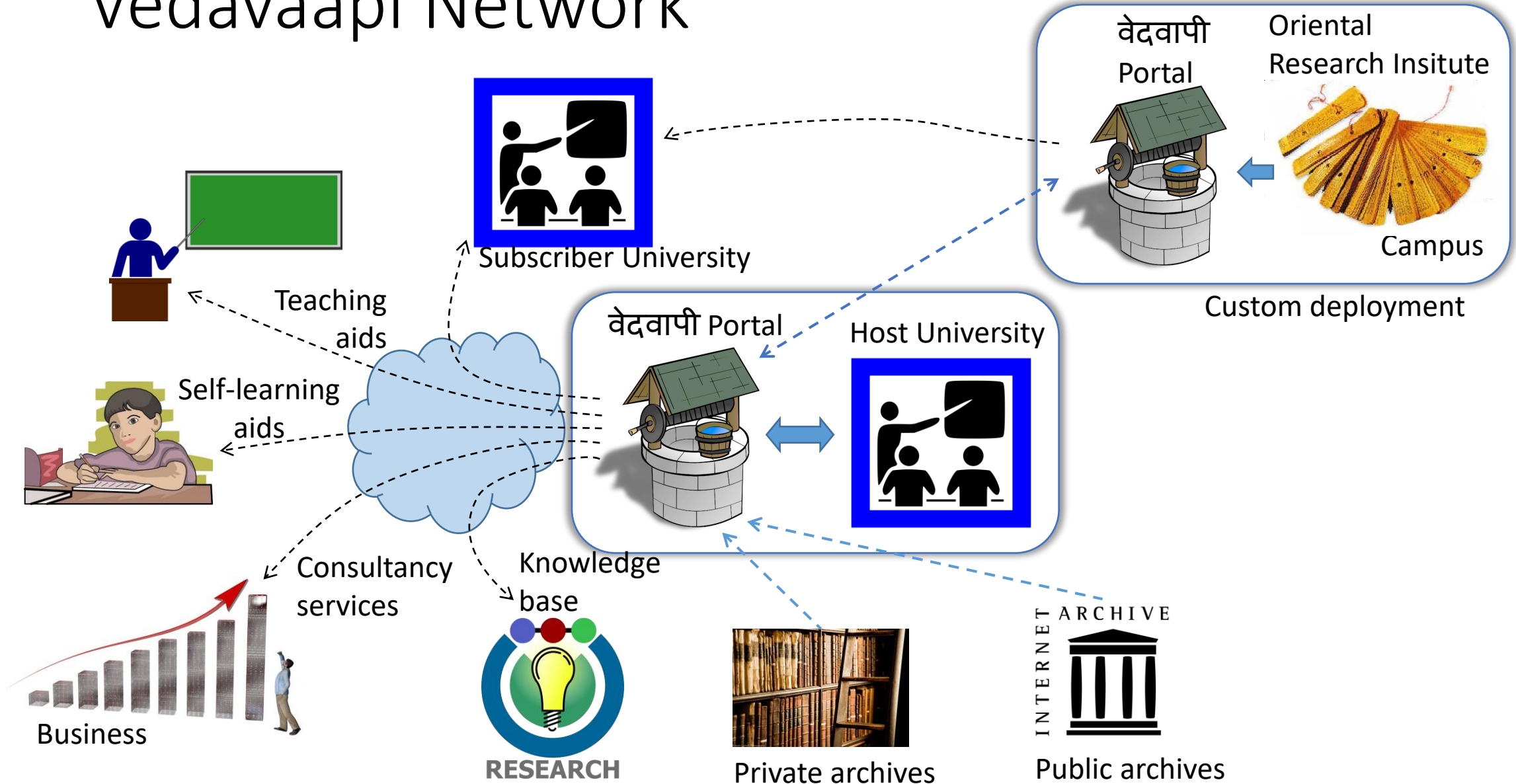
Indic Knowledge: Road to Revival

Innovate for Holistic Living

Explore Bharat's Native Knowledge Base

Achieve Widespread Samskrit Literacy

Vedavaapi focus area

जीवन-शिक्षणम्

शास्त्र-शिक्षणम्

भाषा-शिक्षणम्

Vedavaapi Project Agenda

Uses

Wide sharing, Human-readability

Text Search, Machine-processing

Language assist, Topic search, Concept mining

Comparative study, New insights

Consultancy services, Diagnosis, Recommendations

Tech

Collect → Digitize → Textify → Analyze → Infer → Utilize

Output

- Recordings of experts
- Palm leaf texts
- Scanned Print

- Audio
- Video
- Images
- Text

- Text
- Metadata

- Syntax
- Grammar
- Semantics
- Concepts

- Cross-reference
- Correlations
- Inferences

- Rule base
- Services

Vedavaapi.org
Online Portal

Vedavaapi Network

Oriental Research Insitute

वेदवापी Portal

Campus

Custom deployment

Subscriber University

Teaching aids

Self-learning aids

Consultancy services

Business

वेदवापी Portal

Host University

Knowledge base

RESEARCH

Private archives

INTERNET ARCHIVE

Public archives

Vedavaapi: Technology Areas and Applications

Text Extraction (Media to Text)
- Textract: Indic Manuscript Decoder, Image search tool
- Samskrit Text-to-speech and Speech-to-text tools
- Indexing and search tools for video and audio content

Language Assistance (Text to Concept)
- Samskrit Assistant (grammar, kosha ready-reference) leveraging existing grammar tools
- Grammar-aware Search engine for Samskrit texts

Knowledge Encoding (Concept to Insight)
- Templates to summarize conceptual content of shaastras
- Concept map tools: editing, search, mining, inference

Vedavaapi Initiative: TEXTRACT
Crowd-sourced Manuscript Decoding

- **Goal: Make Heritage Indic documents searchable and processable by blending human and machine intelligence to extract text content**
- Millions of heritage Indic texts are digitized but not amenable to content search, indexing and text processing
- Manuscript deciphering is slow, laborious process (2 years / text)
- Add search and indexing capabilities to digitized Indic content
- Accelerate via crowd-sourcing, image processing, deep learning
- Enables further study of heritage Indic knowledge and research

Vedavaapi Initiative: Shaastra Summary Maps

- **Goal: Help modern seekers with short attention span & low memory training to rapidly grasp and navigate Indic Shaastra content**
- Create a knowledge base + tools to curate it, browse it, process it
- Populate the knowledge base (involve shaastra experts)
  - E.g., encode vaastu rules
- Train in using the knowledge base (students)
  - E.g., help students navigate vaastu text easier
- Create applications leveraging the knowledge base
  - E.g., create a vaastu compliance checker given a building plan

वेदवापी

Experts   Seekers

वेदवापी
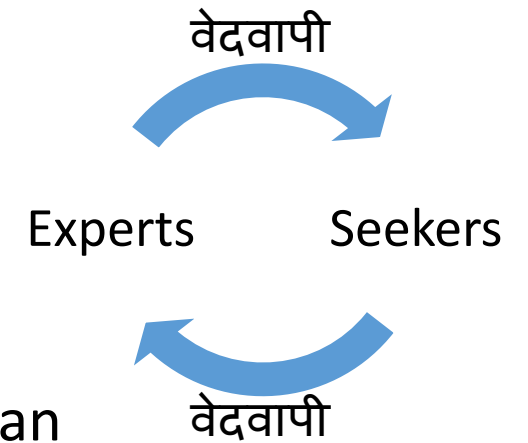
Vedavaapi: Success Criteria

- Vedavaapi used by majority of Indic Science students & scholars
- High shaastra coverage of the Vedavaapi knowledge base
- Popular applications built on Vedavaapi knowledge services

Vedavaapi: Current Status

- Human-assisted OCR: Text Extraction from Scanned Book Images
  - Ready: Web-based editor for crowd-sourced deciphering of scanned texts
  - In progress: Automated text annotation using Deep learning techniques
- Samskrit Language Analysis Service and Programmable (RESTful) API
  - Word analysis: Based on Univ of Hyderabad's Samsaadhanii tools
  - Application: Automated Sentence generation for Samskrit E-learning Exercises
  - Future: Human-assisted Sentence analysis
- Shaastra Summary Maps: Crowd-sourced concept mapping of texts
  - In design stage

Vedavaapi: Initial Funding Requirements

Item	Description	Tenure	Cost (INR)
Server platform (hardware) for 5 years	Amazon EC2 *"t2-medium"* server instance	5 years	100,000 / year * 5 years = **5,00,000**
Software Platform Development (AngularJS web frontend + MongoDB+Python RESTful server backend)	1 web UI developer, 1 backend developer cum architect	1 year	40,000 + 60,000 p.m * 12 = **12,00,000**
Vaakya sangraha (Indic Logic text) concept mapping	Three people + one senior guide	3 months	(30K + 30K + 40K) * 3 = **3,00,000**
Text extraction of a scanned book using Vedavaapi platform (typing, proof-reading) – for Testing Purposes	Two full-time persons with Devanagari data typing experience	6 months	2 * 10000 p.m. * 6 months = **1,20,000**
Total initial fund to launch Vedavaapi service for 5 years			**21,20,000 (21.2 Lakhs) INR or 30,000 USD**

Vedavaapi: How you can Contribute

If you are	Potential Role as contributor
Shaastra Student	User, Content curator, Researcher
Technology Student	Software Intern, Big Data researcher
Shaastra Expert	Content Contributor, Curator, Annotator, Editor
Head of an Institution	Project Host, Paid subscriber of services
Philanthropist	Financial Supporter, fundraiser
IT developer	Platform Software and Application Developer
Enthusiast	Proof-reader, user, tester

Vedavaapi Application: Textract

Rapid Text Extraction from Digitized Indic Manuscripts and Printed Texts

Motivation: State of India's Knowledge Base

- Locked up in 4+ million manuscripts yet to be decoded

- Inaccessible to young students and scholars
  - Hard to decipher and understand (e.g., taalapatras)
  - Steep learning curve, dwindling scholars

- No end-to-end support for collaborative research
  - Deciphering laborious, repetitive, slow (2yrs/text)
  - Disconnected tools, no knowledge-sharing help
 - No cross-reference of texts for comparative study

Textract: Human-assisted OCR for Indic Texts
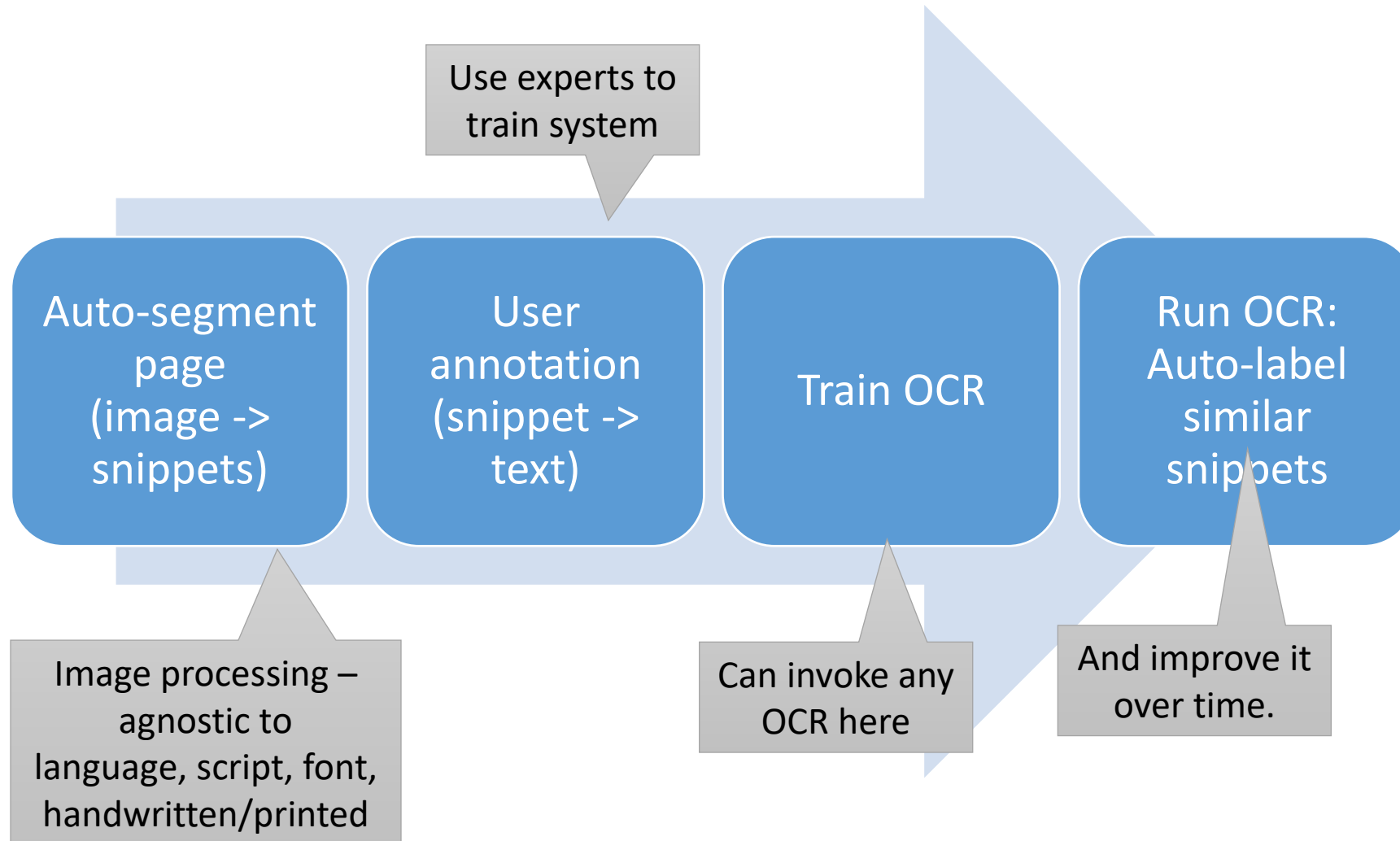
Textract Benefits

- Editors correct mistakes, system re-learns to reduce further mistakes
- Rapid text extraction from scanned books via taking the repetitive work out of proof-reading OCR output
- Language-, script- and font-agnostic text recognition unlike existing OCR
  - Can extract text from printed as well as handwritten text
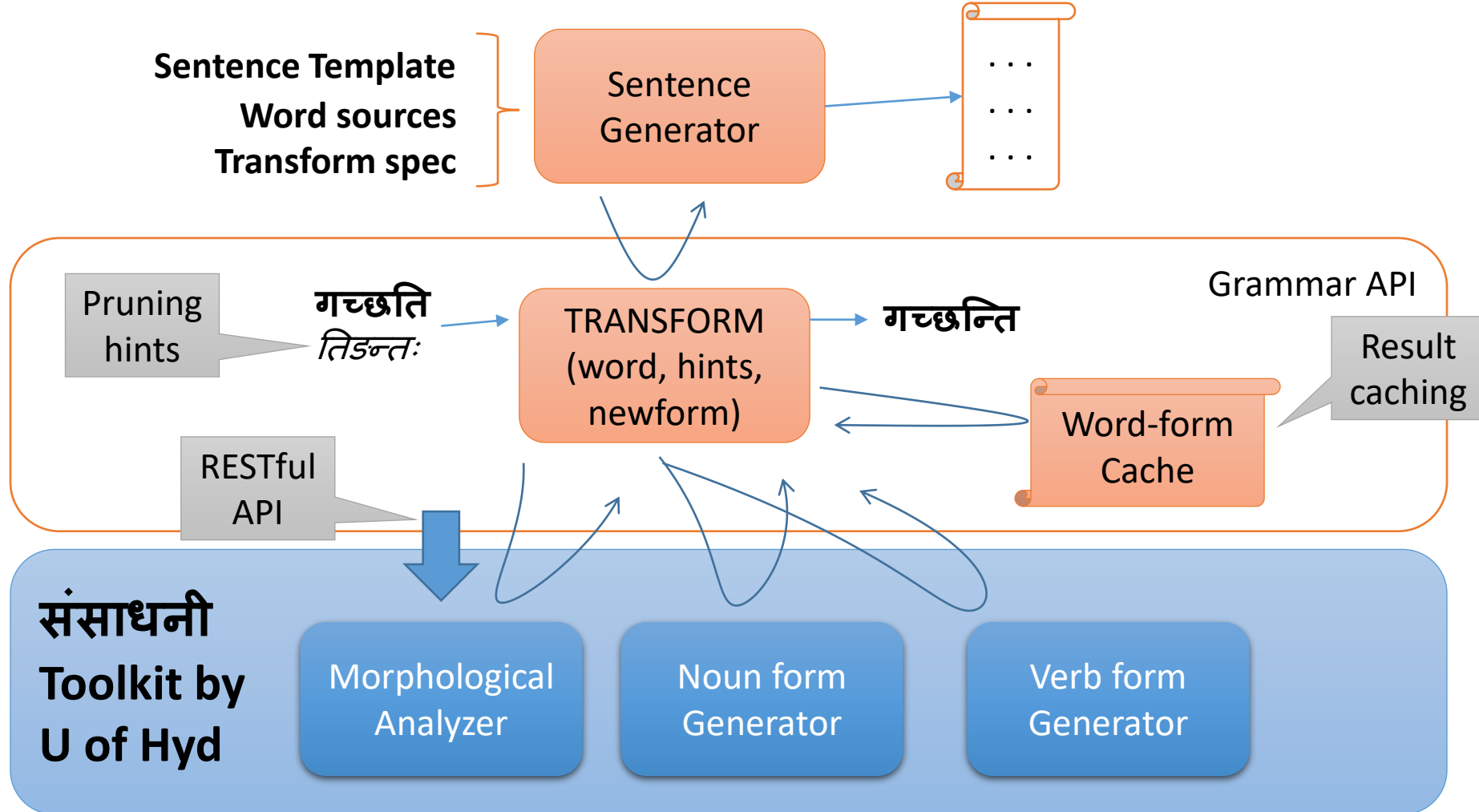
Textract: Example
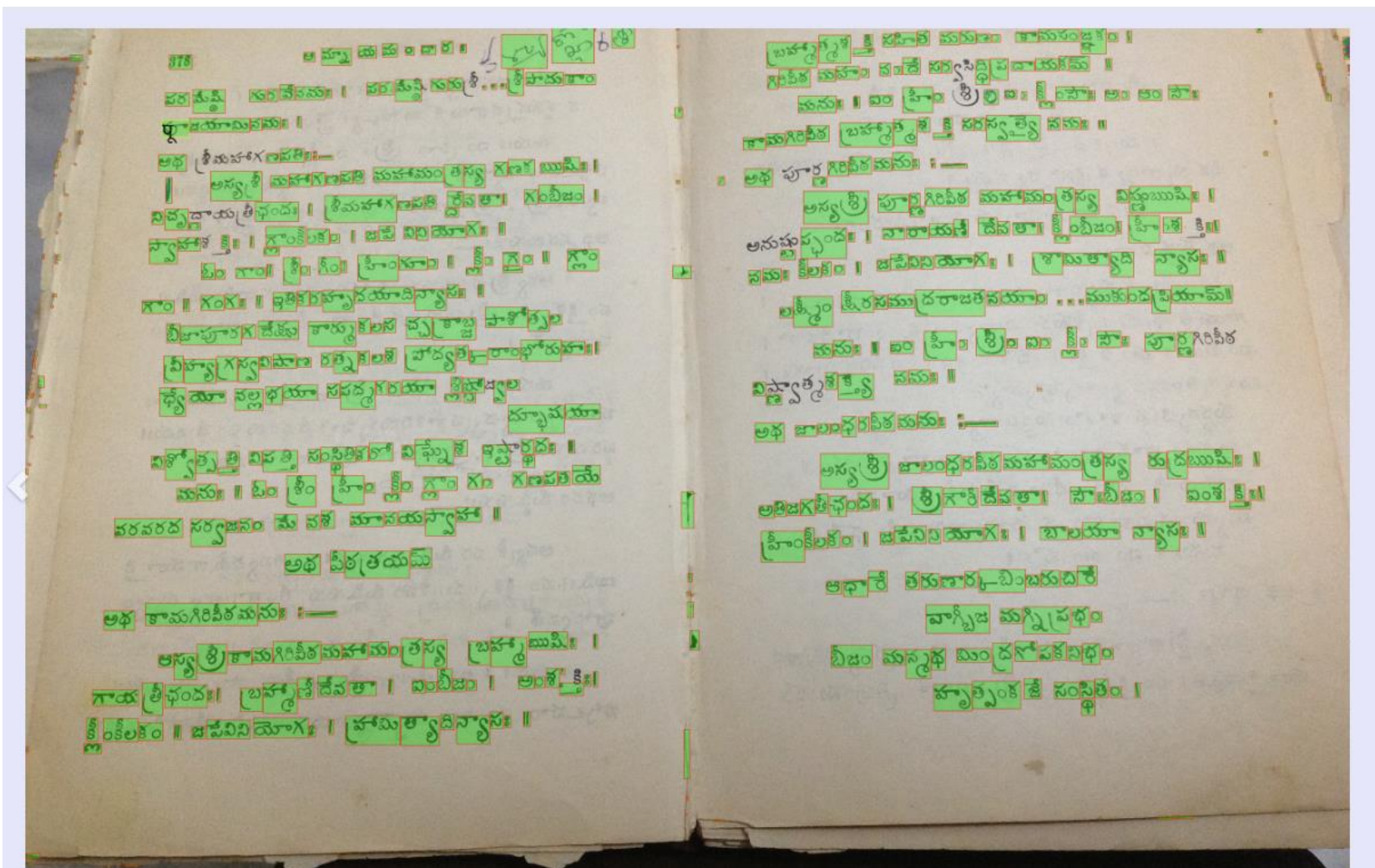
Vedavaapi Application: Samvit

Automated Grammar Exercises for Samskrit Language E-learning

Samvit: Accelerating Samskrit E-learning

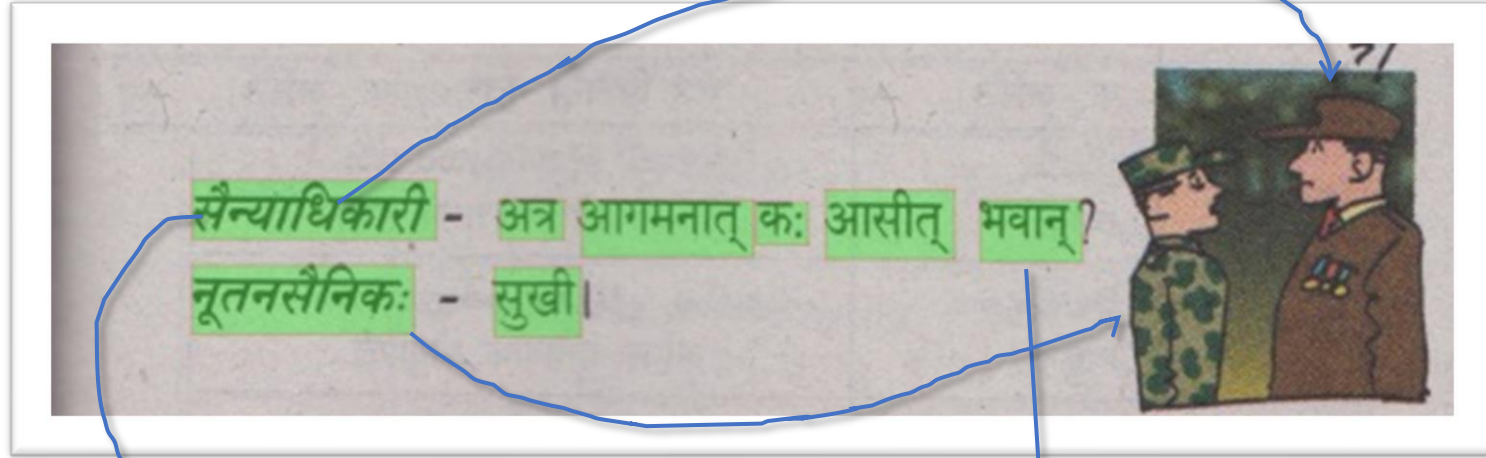[वाहनम्, यानम्, द्विचक्रिका]तृतीया x [गच्छति, याति, चलति]

Vedavaapi Application:

Semi-automated Concept Mapping of Indic Shaastra Texts

Indic Document Analysis Levels: Example

Image

Text

सैन्याधिकारी – अत्र आगमनात् कः आसीत् भवान् ?
नूतनसैनिकः – सुखी ।

Grammar

नूतनः सैनिकः - "सुखी" ।

Semantics

(वदति)   (अहं) सुखी (आसम्) (इति)

Linkages

भवान्   —— Is same as ——→   नूतनसैनिकः

Inference

सुखी   —— Who? ——→ नूतनसैनिकः
       —— When? ——→ अत्र आगमनात् पूर्वम्

- Multiple data formats
- Relationships
- Lots of metadata
- Scale this to 1000's of books x 1000's of pages/book