



## Kunji : A Resource Management System for Higher Productivity in Computer Aided Translation Tools

### Abstract

In this paper, we present Kunji, a resource management system for translation workbenches and MT system development. This system can be easily integrated in translation workbenches and can also be used as a management tool for resources of MT systems. We study the impact of providing this resource management system along with resources, exhibiting linguistic features, on the productivity of translators for English-Hindi language pair. When the resources like lexeme, NER and MWE dictionaries were made available to translators in addition to their regular TM, concordances and terminologies, their productivity increased by 15.61%.

### Kunji:Resource Management System

Kunji is a web based system based on microservices architecture. It can be easily integrated to a translation workbench and provides a mechanism to facilitate translators to use and manage the various types of resources (i.e. like terminologies, glossaries and domain dictionaries along with resources with linguistic features (lexeme dictionaries, NERs, MWEs etc.)). It does not only facilitate the translators but also facilitates the MT modules' development by facilitating language researchers to manage the language engineering resources in a robust manner for various language processing tasks and evaluate them which boosts the process of large scale MT development.

### Features

1. Provision to import different resources in a translation workbench
2. Provision to export the resources and reuse
3. Editing and updating of the resources and their corresponding meta data
4. Addition of the resources in single and bulk mode

5. Search: N-gram dynamic search mechanism
6. Provision of resources with linguistic features - lexemes, NER and MWE dictionaries in addition to their regular dictionaries
7. Provision to verify them with a senior language expert of the corresponding language
8. Easy integration in translation workbench

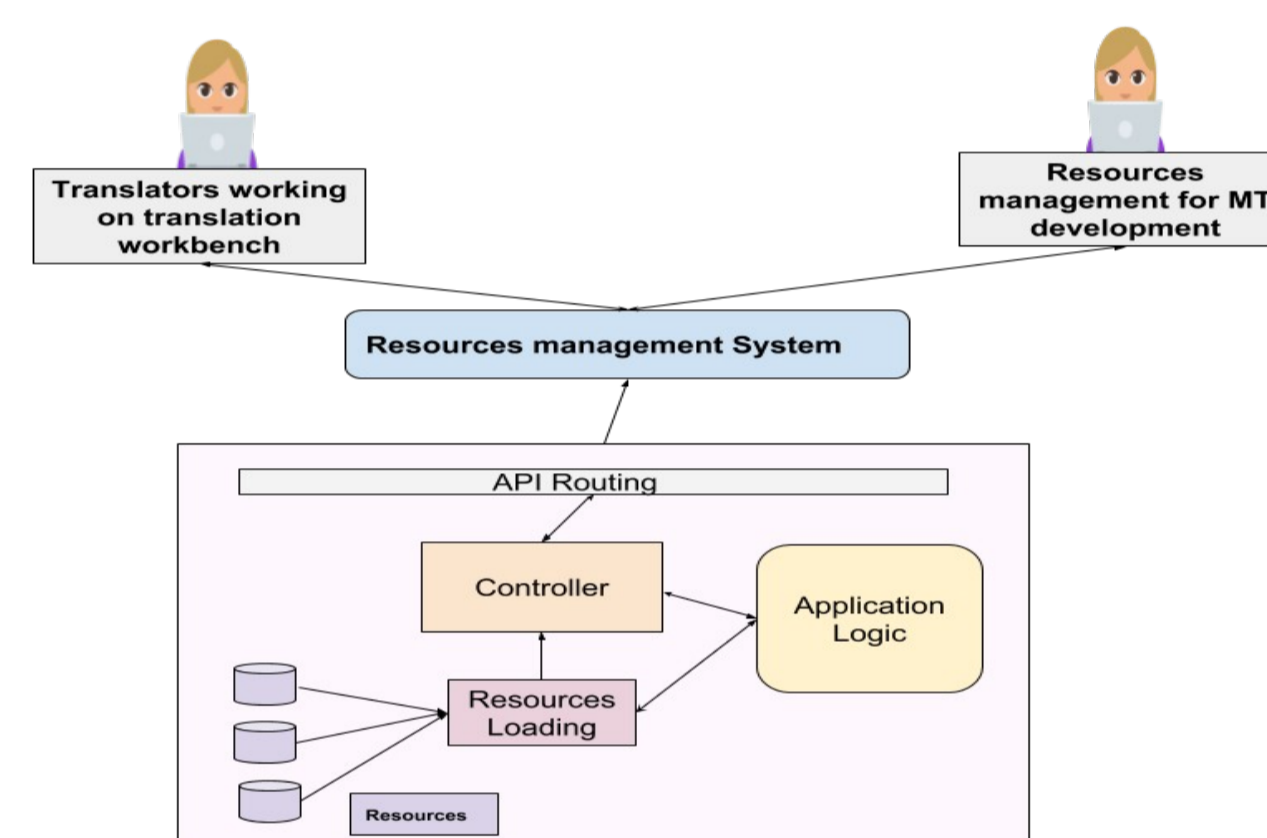


Figure 1 : Kunji Architecture

## EXPERIMENTS AND RESULTS

Data Set of 20 stories from <http://thehindu.com> :

Set No	Paras	Words	Sentences	Avg sentences per para	Avg paras per story	Avg words per para	Avg sentences per story	Avg words per sentence
Set1(5 stories)	42	1180	65	1.54	8.5	28	13	18.15
Set2(5 stories)	64	1722	87	1.36	12.8	26.9	17.4	19.8
Set3(5 stories)	59	1553	83	1.4	11.8	26	16.6	18.71
Set4(5 stories)	54	1238	74	1.37	10.8	23	14.8	16.73
<b>Total</b>	<b>219</b>	<b>5693</b>	<b>309</b>	<b>Avg.</b>	<b>11</b>	<b>26</b>	<b>15.45</b>	<b>18.34</b>

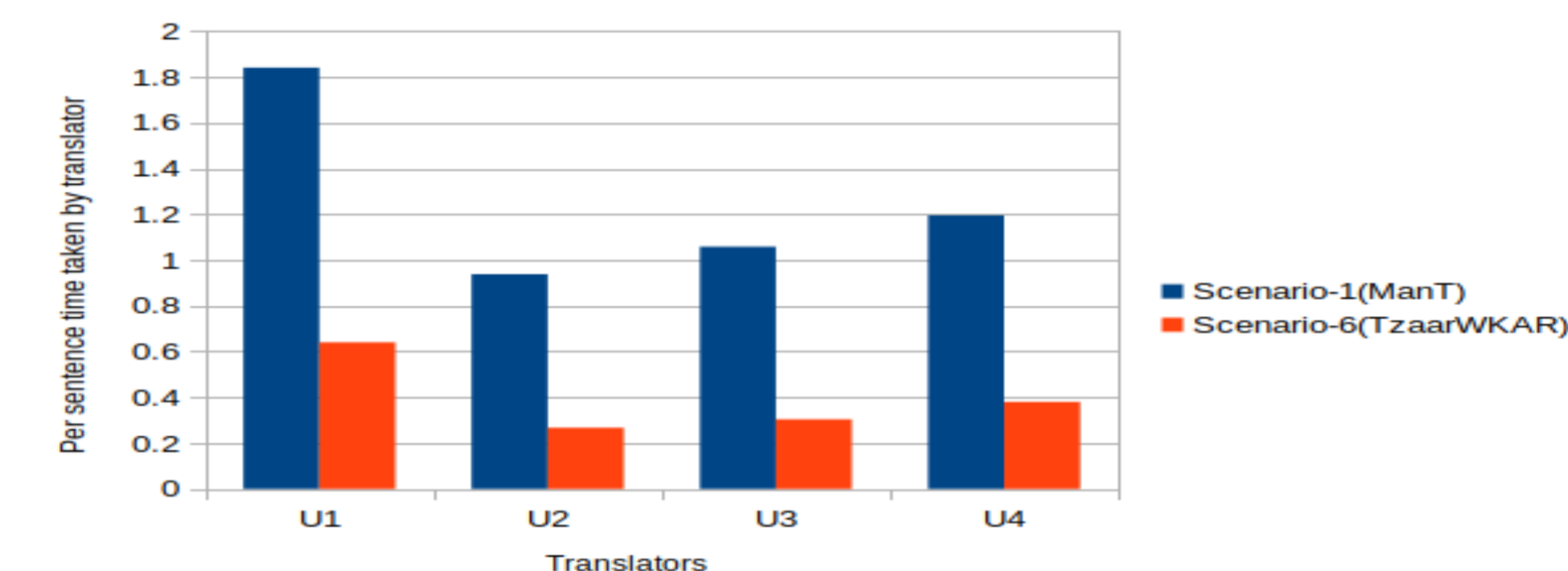
Different Scenarios

Scenarios	Description	Abbr.
Scenario-1	Manual Translation	ManT
Scenario-2	Post-editing in GT on Text editor	PetGT
Scenario-3	GT with Transzaar without Kunji	TzaarWOK
Scenario-4	GT with Transzaar and Kunji (only TM, Terms, Glossaries)	TzaarWKRR
Scenario-5	GT with Transzaar and Kunji (only linguistic resources)	TzaarWKLR
Scenario-6	GT with Transzaar and Kunji (TM, Terms, Glossaries and linguistic resources)	TzaarWKAR

Scenarios	Total Time taken by U1 per sent. (in min.)	Total Time taken by U2 per sent. (in min.)	Total Time taken by U3 per sent. (in min.)	Total Time taken by U4 per sent. (in min.)
Scenario-1(ManT)	1.84(S1)	0.938(S4)	1.058(S3)	1.195(S2)
Scenario-2(PetGT)	0.78(S2)	1.03(S1)	0.569(S4)	0.61(S3)
Scenario-3(TzaarWOK)	0.6(S3)	0.733(S2)	1.07(S1)	0.52(S4)
Scenario-4(TzaarWKRR)	0.506(S2)	0.76(S1)	0.359(S4)	0.43(S3)
Scenario-5(TzaarWKLR)	0.41(S4)	0.49(S3)	0.59(S2)	0.88(S1)
Scenario-6(TzaarWKAR)	0.64(S1)	0.267(S4)	0.304(S3)	0.38(S2)
<b>Improvement (in %)</b>	<b>66%</b>	<b>71.6%</b>	<b>71.3%</b>	<b>68.2%</b>

Scenario-wise productivity of all four translators

Plot of Time Taken by Different Translators in scenario-1 and 6



### Observations

- Productivity of the translators increases as they use the translation workbench with resources
- It increases by more than 66% in Scenario-6 for each of the translators
- Productivity of the translators has been improved additionally by 15.61% over Scenario-4(TzaarWKRR) when they utilize resources with linguistic features along with their regular TM, terminologies and glossaries.
- Set-1 is the toughest and set-4 is easiest of all four sets
- Translators found the dictionaries with linguistic features (lexeme, named entities and MWE ) very helpful

### Conclusion & Future Work

The proposed system addresses the issues of the translators as well as MT system development. Using Kunji, a translator can utilize, reuse and manage his resources which can be shared across projects and can be used further. We hope to extend this system by analyzing the impact of more complex NLP module like Morphological analyzer and Multiword Expressions in the translation pipeline.



## Kunji : A Resource Management System for Higher Productivity in Computer Aided Translation Tools

### Abstract

In this paper, we present Kunji, a resource management system for translation workbenches and MT system development. This system can be easily integrated in translation workbenches and can also be used as a management tool for resources of MT systems. We study the impact of providing this resource management system along with resources, exhibiting linguistic features, on the productivity of translators for English-Hindi language pair. When the resources like lexeme, NER and MWE dictionaries were made available to translators in addition to their regular TM, concordances and terminologies, their productivity increased by 15.61%.

### Kunji:Resource Management System

Kunji is a web based system based on microservices architecture. It can be easily integrated to a translation workbench and provides a mechanism to facilitate translators to use and manage the various types of resources (i.e. like terminologies, glossaries and domain dictionaries along with resources with linguistic features (lexeme dictionaries, NERs, MWEs etc.)). It does not only facilitate the translators but also facilitates the MT modules' development by facilitating language researchers to manage the language engineering resources in a robust manner for various language processing tasks and evaluate them which boosts the process of large scale MT development.

### Features

1. Provision to import different resources in a translation workbench
2. Provision to export the resources and reuse
3. Editing and updating of the resources and their corresponding meta data
4. Addition of the resources in single and bulk mode

5. Search: N-gram dynamic search mechanism
6. Provision of resources with linguistic features - lexemes, NER and MWE dictionaries in addition to their regular dictionaries
7. Provision to verify them with a senior language expert of the corresponding language
8. Easy integration in translation workbench

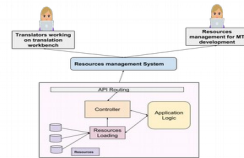


Figure 1 : Kunji Architecture

## EXPERIMENTS AND RESULTS

Data Set of 20 stories from <http://thehindu.com> :

Set No	Para	Words	Sentences	Avg sentences per para	Avg words per para	Avg words per sentence
Set1 (manus)	42	1140	65	1.52	17.5	13.7
Set2 (manus)	64	1722	87	1.36	12.8	26.9
Set3 (manus)	99	1453	83	1.4	11.8	26
Set4 (manus)	54	1238	74	1.37	10.8	23
Total	259	5693	309	Avg. 11	26	15.45

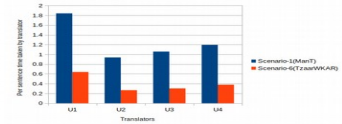
### Different Scenarios

Scenarios	Description	Abbr.
Scenario-1	Manual Translation	ManT
Scenario-2	Post-editing in GT on text editor	PEGT
Scenario-3	GT with Transzear without Kunji	TzearWOK
Scenario-4	GT with Transzear and Kunji (only TM, Terms, Glossaries)	TzearWKRR
Scenario-5	GT with Transzear and Kunji (only linguistic resources)	TzearWKLR
Scenario-6	GT with Transzear and Kunji (TM, Terms, Glossaries and linguistic resources)	TzearWKAR

Scenarios	Total Time taken by U1 per sent. (in min.)	Total Time taken by U2 per sent. (in min.)	Total Time taken by U3 per sent. (in min.)	Total Time taken by U4 per sent. (in min.)
Scenario-1(ManT)	1.84(S1)	0.93(S2)	1.68(S3)	1.19(S4)
Scenario-2(PEGT)	0.78(S1)	0.03(S2)	0.50(S3)	0.01(S4)
Scenario-3(TzearWOK)	0.63(S1)	0.73(S2)	1.07(S3)	0.23(S4)
Scenario-4(TzearWKRR)	0.50(S2)	0.76(S1)	0.39(S4)	0.43(S3)
Scenario-5(TzearWKLR)	0.41(S4)	0.49(S3)	0.59(S2)	0.88(S1)
Scenario-6(TzearWKAR)	0.58(S1)	0.26(S4)	0.34(S3)	0.38(S2)
Improvement (in %)	66%	71.6%	71.3%	68.2%

Scenario-wise productivity of all four translators

Plot of Time Taken by Different Translators in scenario-1 and 6



### Observations

- Productivity of the translators increases as they use the translation workbench with resources
- It increases by more than 66% in Scenario-6 for each of the translators
- Productivity of the translators has been improved additionally by 15.61% over Scenario-4(TzearWKRR) when they utilize resources with linguistic features along with their regular TM, terminologies and glossaries.
- Set-1 is the toughest and set-4 is easiest of all four sets
- Translators found the dictionaries with linguistic features (lexeme, named entities and MWE ) very helpful

### Conclusion & Future Work

The proposed system addresses the issues of the translators as well as MT system development. Using Kunji, a translator can utilize, reuse and manage his resources which can be shared across projects and can be used further. We hope to extend this system by analyzing the impact of more complex NLP module like Morphological analyzer and Multiword Expressions in the translation pipeline.