

KantanMT.com No Hardware. No Software. No Hassle MT.

Lexically Constrained Decoding for Sequence Generation Approaches to support Terms, Tags & Placeholders

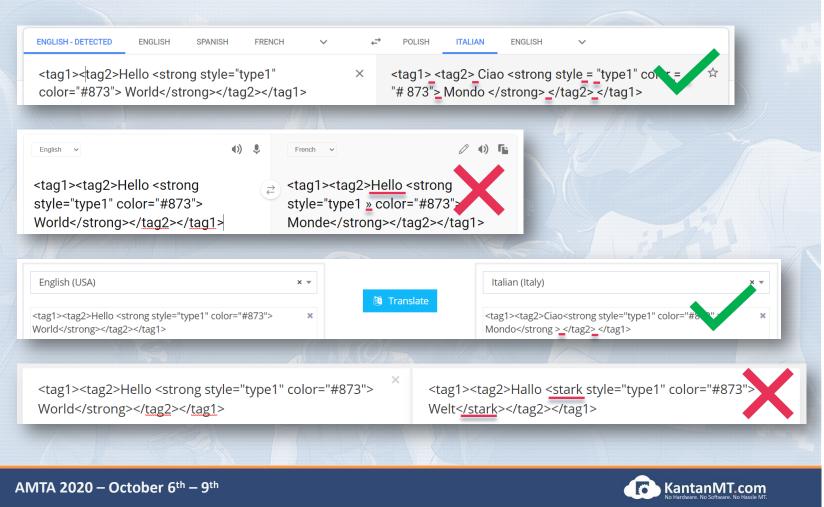
The Problem....

original	Hello <g ctype="x-bold;" id="1">World!</g>				
after tokenization	Hello < g id = " 1 " ctype = " x-bold ; " > World ! < / g >				
after escaping	Hello < g id = " 1 " ctype = " x-bold ; "				
	> World ! < / g >				
after lowercasing	hello < g id = " 1 " ctype = " x-bold ; "				
	> world ! < / g >				
after translation	guten tag < code < ept id = " 1 "				
	ctype = " " " x-bold ; " > world money				
	< / g > !				
after recasing	Guten Tag < Code & lt; ept id = & quot; 1 & quot;				
	ctype = " " " x-bold ; " > World Money				
	< / G > !				
after detokenization	n Guten Tag < Code < ept id = " 1				
	<pre>&quot; ctype = "" "x-bold;" > World Money &lt; / G</pre>				
	>!				
after unescaping	Guten Tag < Code < ept id = " 1 " ctype = ""				
	"x-bold;" > World Money < / G >!				

AMTA 2020 – October 6th – 9th



The Results...



Subtitling DFXP Format

And if that is not enough world championship action,
you can pick up the World Championship Pikachu Poké Plush
 begin="00:01:21:19" end="00:01:23:02">at pokémoncenter.com.





Proceedings of the 14th Conference of the Association for Machine Translation in the Americas October 6 - 9, 2020, Volume 2: MT User Track

Subtitling DFXP Format

And if that is not enough world championship action,
you can pick up the World Championship Pikachu Poké Plush
at pokémoncenter.com.

Step 1 - Sentence Forming

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush pokémoncenter.com.





Subtitling DFXP Format

And if that is not enough world championship action,
you can pick up the World Championship Pikachu Poké Plush
at pokémoncenter.com.

Step 1 - Sentence Forming

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush at pokémoncenter.com.

Step 2 - Translate

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship Pikachu Poké Plush na pokémoncenter.com.





Step 1 - Sentence Forming

And if that is not enough world championship action, you can pick up the World Championship Pikachu Poké Plush at pokémoncenter.com.

Step 2 – Translate

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship Pikachu Poké Plush na pokémoncenter.com.

Step 3 – Reformat & Re-insert

И если этого недостаточно для участия в чемпионате мира,
вы можете забрать Pikachu Poké Plush Plush
 begin="00:01:21:19" end="00:01:23:02">pokémoncenter.com.





Strategy #2: Zones & Walls

Moses decoder has the concept of

 Forced Translations – parts of the input sentence can be wrapped in <np> tags and assigned a probability

<np translations="Translated Term" prob="1">...</np>

- Zones
 - Encapsulates a series of tokens that cannot be individually reordered, but can be re-ordered as part of the parent sequence

You can pick up the World Championship <zone>Pikachu Poké Plush</zone>

- Wall
 - A hard border that tokens cannot cross during re-ordering

AMTA 2020 – October 6th – 9th



Strategy #3: Segmentation

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.



AMTA 2020 – October 6th – 9th



Strategy #3: Segmentation

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship

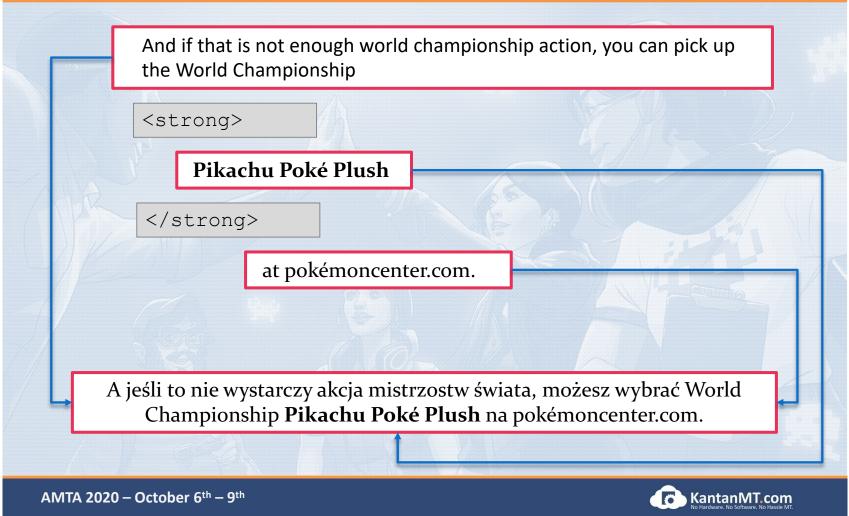
Pikachu Poké Plush

at pokémoncenter.com.

AMTA 2020 – October 6th – 9th



Strategy #3: Segmentation



And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

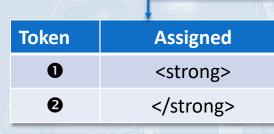


AMTA 2020 – October 6th – 9th



And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship **●** Pikachu Poké Plush **②** at pokémoncenter.com.

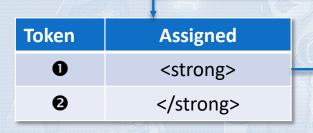


AMTA 2020 - October 6th - 9th



And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship **1** Pikachu Poké Plush **2** at pokémoncenter.com.



A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **OPikachu Poké Plush** na pokémoncenter.com.

AMTA 2020 - October 6th - 9th



And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

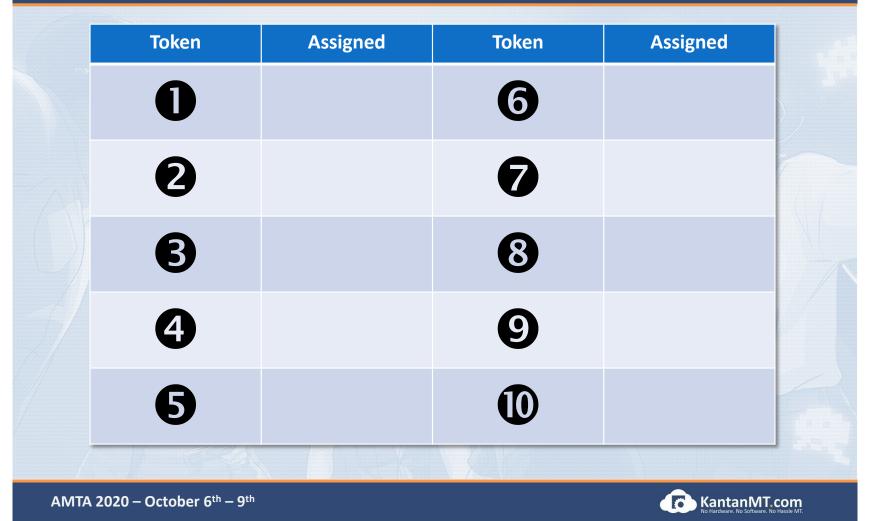
And if that is not enough world championship action, you can pick up the World Championship **●** Pikachu Poké Plush **②** at pokémoncenter.com.



A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na pokémoncenter.com.







Strategy #5: Optimisation

Detecting lead/trail tags

<font="Calibri"><color="#1234"><size="10pts">And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com. </size></color>



AMTA 2020 – October 6th – 9th



Strategy #5: Optimisation

Detecting lead/trail tags

<font="Calibri"><color="#1234"><size="10pts">And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com. </size></color>

<lead/>And if that is not enough world championship action, you can pick up the World Championship **1** Pikachu Poké Plush **2** at pokémoncenter.com.<trail/>

	Token	Assigned
	0	
-	0	
	Lead	<font="calibri"><color="#1234"><size="10pts"></size="10pts"></color="#1234"></font="calibri">
	Trail	

AMTA 2020 - October 6th - 9th

KantanMT.com

Which one to work with???

- Strategy #1
 - Structure Forming
- Strategy #2
 - Zones & Walls
- Strategy #3
 - Segmentation
- Strategy #4
 - Pass-Thru
- Strategy #5
 - Optimisation







KantanMT.com No Hardware. No Software. No Hassle MT.

Challenges of Terminology

Basic Levels of Terminology

Lexically Constrained Decoding for Sequence Generation

- Level I
 - Multi-word phrases non-inflected
 - Nouns, Proper Nouns, Brand or Feature/Service names
 - Can be represented in simple list form

EN	FR	IT	DE	
Stakataka	Ama-Ama	Stakataka	Muramura	
Snorlax	Ronflex	Snorlax	Relaxo	
Trainer	Dresseur	Allenatore	Trainer	
Terrakion	Terrakium	Terrakion	Terrakium	
Whimsicott	Farfaduvet	Whimsicott	Elfun	
Gengar	Ectoplasma	Gengar	Gengar	
Tapu Bulu	Tokotoro	Tapu Bulu	Kapu-Toro	
Kommo-o	Ékaïser	Kommo-o	Grandiras	

• Level II

- Inflected Multi-word phases
- Cannot be represented in simple list form
- Inflected variations to indicate number, grammatical case, or gender

AMTA 2020 - October 6th - 9th



What is a Glossary?

Terms in one or more languages

- With comments, and usage examples
- Industry Glossary
 - Includes terms that are standard for an industry or subject/domain
 - **Example**: Banking, Cardiology, Automotive, Accounting
- Client Glossary
 - Contains terms that are very specific to a company/client
 - Example: eBay, Adobe, BMW
- Project Glossary
 - Used to maintain consistency throughout translation project
 - Example: Microsoft Office, Pokemon



https://iate.europa.eu/home

Terms: 7.9 million Languages: 26



AMTA 2020 – October 6th – 9th

Level #1: Segmentation

And if that is not enough world championship action, you can pick up the World Championship <term>**Pikachu Poké Plush** </term>at pokémoncenter.com.

And if that is not enough world championship action, you can pick up the World Championship

<term>

Pikachu Poké Plush

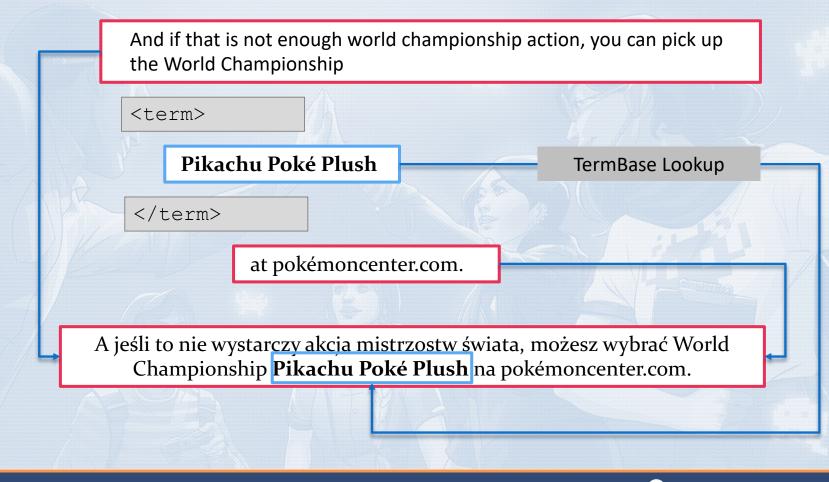
</term>

AMTA 2020 – October 6th – 9th

at pokémoncenter.com.



Level #1: Segmentation



AMTA 2020 – October 6th – 9th



Level #1: Pre-process & Pass-Thru

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

TermBase Lookup And if that is not enough world championship action, you can pick up the World Championship <**np translations="Pikachu Poké Plush" prob="1"**/> at pokémoncenter.com.



Level #1: Pre-process & Pass-Thru

And if that is not enough world championship action, you can pick up the World Championship **Pikachu Poké Plush** at pokémoncenter.com.

TermBase Lookup And if that is not enough world championship action, you can pick up the World Championship <**np translations="Pikachu Poké Plush" prob="1"**/> at pokémoncenter.com.

A jeśli to nie wystarczy akcja mistrzostw świata, możesz wybrać World Championship **Pikachu Poké Plush** na pokémoncenter.com.

AMTA 2020 - October 6th - 9th





KantanMT.com No Hardware. No Software. No Hassle MT.

Thank you...