

State-of-art technology in NLP / NLU / NLG

Understanding Natural Language the way native speakers do

The problem

For years traditional approaches to Natural Language Processing (NLP) relied on keyword occurrences, underestimating the need for specific target-language-oriented syntax and grammar to understand real texts

Can a computer tell what is a **real fact** and what is **not a real fact**?

several presentations for the same person (identity matching)

A Mauricio Macri le gusta que el servicio de **Telefónica haya mejorado**.

Le gusta que el servicio de **Telefónica no haya mejorado**.

No le gusta que el servicio de **Telefónica haya mejorado**.

No es verdad que **Telefónica haya mejorado**.

El Ing. Macri exige que el servicio de **Telefónica mejore**.

El presidente Macri espera que el servicio de **Telefónica mejore**.

Él preguntó si **Telefónica mejoró el servicio de la empresa Movistar**.

Telefónica mejoró el servicio de la empresa Movistar.

Desmintieron el rumor que había sido difundido por el ingeniero Macri acerca de que **Telefónica mejoró el servicio de la empresa Movistar**.

Our solution: SentiLecto

Understanding Natural Language the way native speakers do... <http://natural.do>

- ✓ [SentiLecto](#) is our NLU engine for Spanish and Portuguese. This **Natural Language Understanding (NLU)** solution yields a highly fine-grained representation for the texts. [SentiLecto](#) leans on outstanding linguistic features such as: passive/active voice transformation, negation scope, anaphora resolution and co-reference chains, ellipsis, modality treatment, semantic features and accurate verbal frames for all verbs, even with 'se-impersonal' usages.
- ✓ It can recognize and classify named-entities (NERC) such as Places, Organizations and Persons with identity matching. Also, [SentiLecto](#) can **identify whether or not an utterance is a real fact (fact mining)**, normalizing facts through deep understanding of syntax and semantics.
- ✓ Finally, [SentiLecto](#) better represents the premise whereby the entities involved are syntactically mapped onto SVO (subject-verb-object) slots for their differential sentiment analysis: '*Mary hates John*' (two entities but only the direct object has a negative presentation) vs. '*Mary defames John*' (the same two entities but only the subject has negative presentation).

on-line demo: <http://dev.natural.do/sentilecto>

How does our NLU technology work?

Overall text score (+0.00% , -79.35%)



Check [Entity-based Sentiment Overall Report](#) at the bottom of this page

Sentence 1

Antônio Silveira, último presidente da FIFA, bateu as botas em 2005. (sentence score 0.00,-72.13)

Clause 1.1 (mouse over SVO slots, entities and sentiment scores for specific attributes to appear)

Antônio_Silveira , último presidente de a FIFA , bateu_as_botas em 2005 . (clause score 0.00,-72.13)

Opinionated Entities (sentiment scores)

	Positive	Negative	Named-Entity
Antônio_Silveira (apposition: último presidente de a FIFA)	0	50	PERSON

Global operators (any checked flag indicates the utterance mapped onto the following SVO slots is not a real fact but an hypothetical situation that didn't happen)

- Negation Yes/No Question Possibility Condition Imperative Future Volitive

Syntactic Structures (SVO slots)

Opinionated Modal Adverb	Subject	Verb	Direct Object	Indirect Object Experiencer	Phrasal-Verb Complement	Equative	Instrument/Company Goal/Beneficiary	Adjuncts (Space, etc.)	Time
	Antônio_Silveira (apposition: último presidente de a FIFA)	bateu_as_botas							em 2005

Normalized fact: antônio_silveira + bateu_as_botas --TIME: em 2005

Sentence 2

Silveira e Renata Souza foram condenados pela policia. (sentence score 0.00,-86.56)

Clause 2.1 (mouse over SVO slots, entities and sentiment scores for specific attributes to appear)

Silveira e Renata_Souza foram condenados por a policia . (clause score 0.00,-86.56)

Opinionated Entities (sentiment scores)

How does our NLU technology work?

Fact extraction

tecnolecto.com/sentilecto/result?text=Cancillería+condenó+hoy+..."

Named-Entities & Fact extraction

- Paul Singer
 - Occurrences (1)
 - Facts
 - cuyo propietario es el especulador internacional Paul_Singer
 - Normalized Facts
- Congreso de Estados Unidos
- Rio_Negro
- Argentina
- Patagonia_Sur
- la República_Argentina
- Cancillería
- Estados_Unidos
 - Occurrences (2)
 - Facts
 - acusó de llevar_a_cabo presiones ante congresistas de Estados_Unidos .
 - Los buitres acusan falsamente a la Argentina de perseguir el objetivo de min...
 - Normalized Facts
- Patagonia Norte B
- Relaciones Exteriores
- Cancillería
 - Occurrences (3)
 - (tacit) él/ella (Cancillería) (2)
 - Cancillería (1)
 - Facts
 - Cancillería condenó hoy - la campaña de hostigamiento de los fondos buitres
 - acusó de llevar_a_cabo presiones ante congresistas de Estados_Unidos .
 - el que criticó
 - Normalized Facts
- el fondo NML

Usage:

El Gobierno adujo la negat... x +

www.lanacion.com.ar/1709099-el-gobie

Viernes 11 de julio de 2014 | 18:05

El Gobierno adujo la negativa a las exportaciones de carne en EE.UU. a "una campaña de los fondos buitre"

Cancillería acusó a los holdouts de hacer "lobby" para que se impida la apertura a la Argentina del mercado en los Estados Unidos

Comentá 898 Facebook 257 Twitter 89 +



Cancillería condenó hoy "la campaña de hostigamiento de los fondos buitres" a cuyo "lobby" acusó de llevar a cabo presiones ante congresistas de Estados Unidos para que impidan la apertura de las exportaciones argentinas de carne a ese país.

"Los buitres acusan falsamente a la Argentina de perseguir el objetivo de minar los estándares de seguridad y salud animal en Estados Unidos", denunció la Cancillería en un comunicado, en el que criticó "en especial" al fondo NML, "cuyo propietario es el especulador internacional Paul Singer".

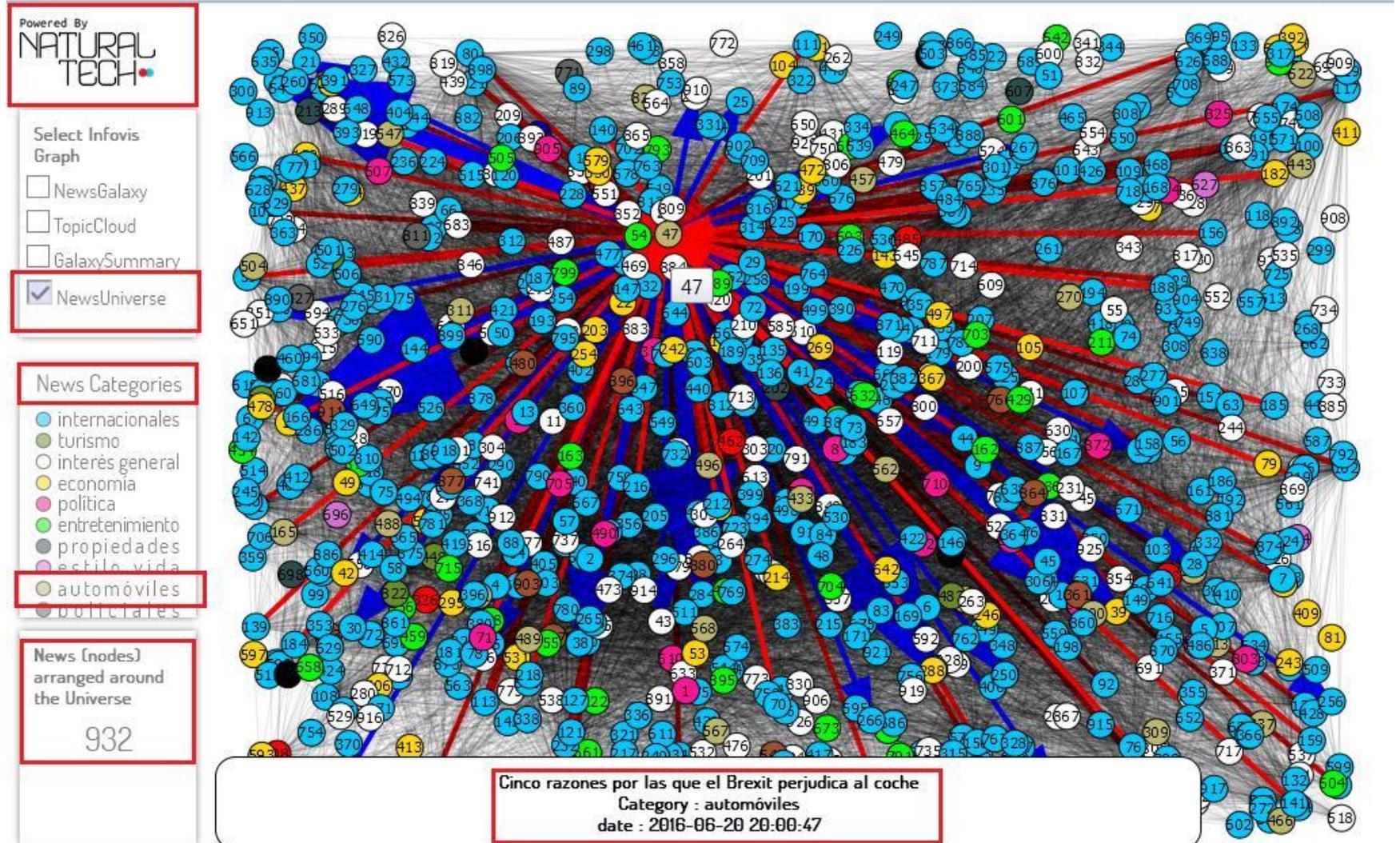
Ese fondo, sostuvo Relaciones Exteriores, "ha ampliado sus agresivas acciones al atacar el proceso productivo y la capacidad exportadora de la República Argentina". "Los lobbys de los fondos buitres realizaron ayer (por el jueves) una presentación en el Congreso de Estados Unidos a fin de presionar políticamente para que se impida la apertura de las exportaciones argentinas de carne a dicho país", indicó.

Entidades mais mencionadas e sua valorização na notícia:

Id	Entity	Positive	Negative	Named-Entity	Total occurrences	Occurrences (appearances)
1	Michel Temer	60	0	PERSON	6	Miche_Temer: 1, Temer: 1, (tacit) ele/e/a (referent: o vice-presidente Michel_Temer): 2, o vice-presidente Michel_Temer: 2
2	o governo	0	0	NONE	5	(tacit) ele/e/a (referent: um advogado): 4, (tacit) ele/e/a (referent: o governo): 1
3	eu	0	0	NONE	3	(tacit) eu: 3
4	golpe	0	0	NONE	3	um golpe: 2, golpe: 1
5	nota	0	0	NONE	3	a nota: 1, nota: 2
6	partido	60	160	NONE	2	o partido liderado: 1, um partido único: 1
7	ato	60	0	NONE	2	ações populares: 1, um ato: 1
8	pmdb	40	0	ORGANIZATION	2	o pmdb: 1, (tacit) ele/e/a (referent: o pmdb): 1
9	Lula	0	0	PERSON	2	Lula: 1, (tacit) ele/e/a (referent: Lula): 1
10	Renan Calheiros	0	0	PERSON	2	Renan_Calheiros: 1, (tacit) ele/e/a (referent: Renan_Calheiros): 1

How does our NLU technology work?

How does our NLU technology work?



What can our NLU technology be potentially used for?

A billionaire market is eagerly waiting for powerful NLU frameworks...

- ✓ Virtual assistants (question & answer systems)
- ✓ Buzz monitoring on social media (sentiment analysis)
- ✓ Machine Translation (grammar representation)
- ✓ Customization of digital contents (verticals and geolocations)

What do we actually use our NLU technology for?

- ✓ Media monitoring (understanding news)
- ✓ Visual Insights for big data based on text datasets
- ✓ Summarization and Topic Detection
- ✓ Natural Language Generation (NLG) (for SEO & SEM campaigns)

How does our NLG architecture work?

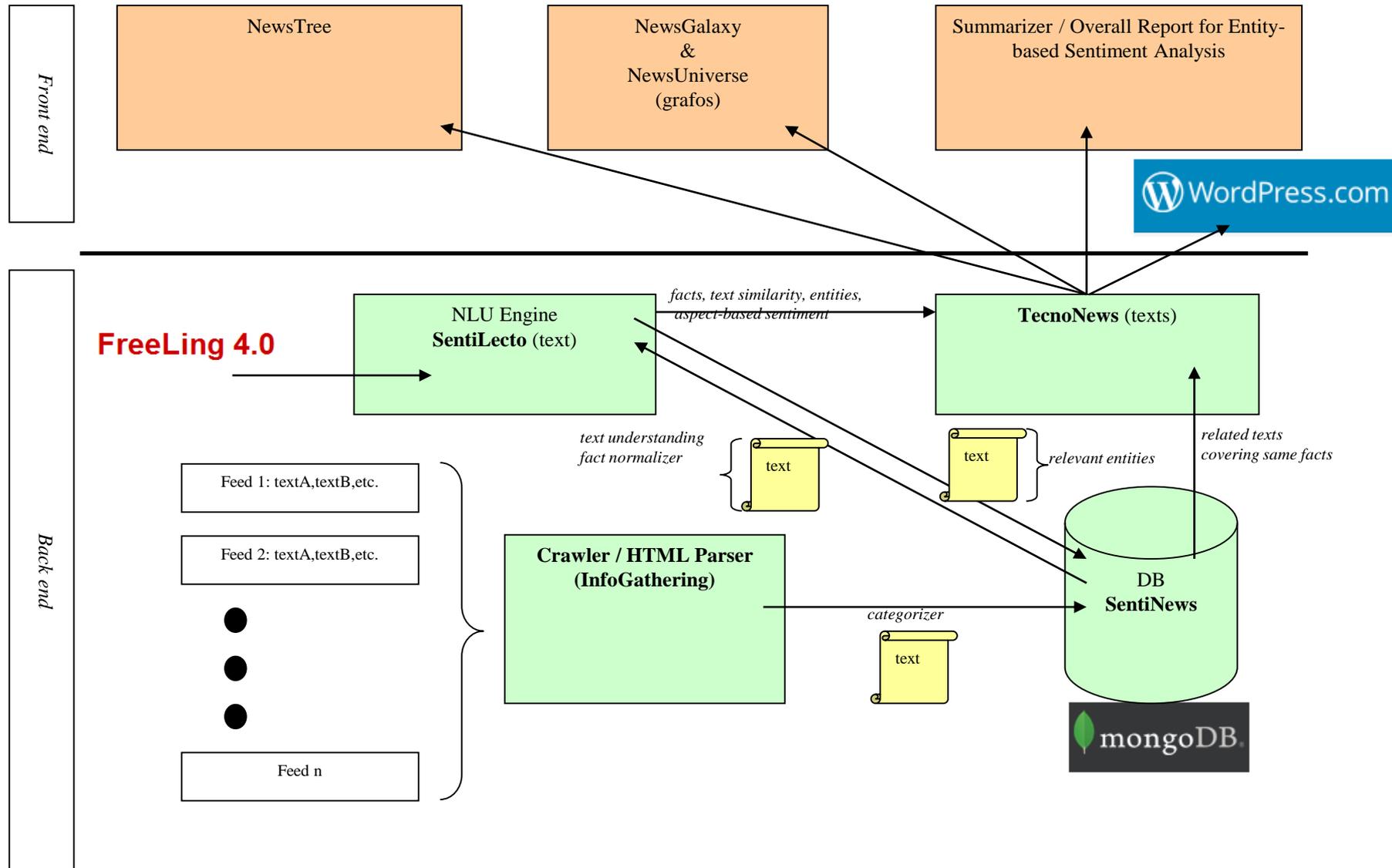
SentiLecto understands, rewrite and merge news all over the world to generate quality posts in specific verticals.

Fine-grained relationships between news covering the same facts and powerful linguistic transformations allow SentiLecto to generate quality posts more informative than the very sources they are based on, as it can identify and merge missing facts or even come up with racconto stories from the scratch.

SentiLecto is being used to automatically generate this blog <http://entretenimientobit.com> with more than 300 quality posts (including images) from 150 feeds on a daily basis.

How does our NLG architecture work?

Showcase <http://entreenimentobit.com>



How does our NLG architecture work?

A história desta notícia a partir de notícias prévias:

- Em resposta a Lula, Temer declara que por ser expert sabe que não há golpe
 - Lula diz ver "com certa tristeza" eventual saída do PMDB do governo - March 28, 2016 (EntretenimientoBit)
 - Governo Dilma usa cargos para atrair siglas da base - March 28, 2016 (EntretenimientoBit)
 - PMDB-RJ sinaliza com desembarque do governo após ofensiva de Dilma - (FolhaGeneric)
 - Governo tenta esvaziar reunião do PMDB que definirá desembarque - (FolhaGeneric)
 - 'Coalizão' com parte do PMDB pode ajudar a barrar impeachment, diz Lula - March 28, 2016 (EntretenimientoBit)
 - Mujica sai em defesa de Dilma e Lula e alerta para tentativa de golpe - (EfeGeneric)
 - OAB entrega na segunda novo pedido de impeachment de Dilma - March 23, 2016 (EntretenimientoBit)
 - OAB declara apoio ao impeachment da presidente Dilma - (FolhaGeneric)
 - Em posse de Lula, Dilma faz duras críticas a Moro e à divulgação de gravações - March 19, 2016 (EntretenimientoBit)
 - Bovespa fecha em queda de 1% com ruídos políticos envolvendo Lula - March 14, 2016 (EntretenimientoBit)
 - Bovespa recua 1,55% com ruído sobre Lula em ministério e queda de commodities - (Extraoglobo-pt)
 - Lula deve aceitar convite de Dilma para ser ministro - March 14, 2016 (EntretenimientoBit)
 - Manifestantes protestam contra Dilma e Lula na Avenida Paulista - (EfeGeneric)
 - É absurdo achar que STF é menos capaz de investigar Lula, diz defesa - (FolhaGeneric)
 - Pessimista, Lula reconhece iminência de derrota, afirmam interlocutores - (FolhaGeneric)
 - Dilma e Lula decidem aumentar espaço de PP e PR no ministério - March 30, 2016 (EntretenimientoBit)
 - Com ruptura, ministros do PMDB se preparam para deixar governo Dilma - March 29, 2016 (EntretenimientoBit)
 - PMDB decide esta semana sobre permanência no governo - (Exametecnologia-pt)

How does our NLG architecture work?

OUR POST
More informative than
original

Item 1
176 words, 30% matched

Item 2
74 words, 70% matched

POST
ORIGINAL

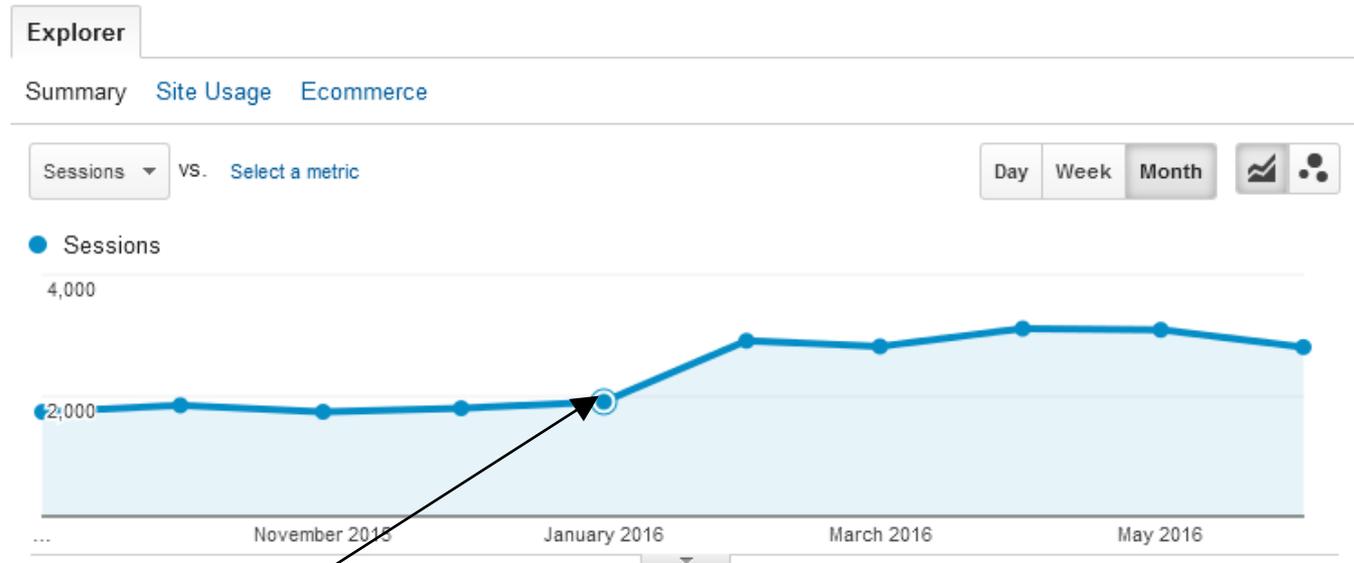
Ex-panicat se divertiu ao lado de Tatá Werneck e Fábio Porchat Na tarde desta terça-feira, <u>Nicole Bahls</u>	« 20 words »	Ex-panicat se divertiu ao lado de Tatá Werneck e Fábio Porchat Na tarde desta terça-feira, Nicole Bahls
<u>que vai estar</u> no programa Ferdinando Show,		
esteve nos estúdios do Multishow para gravar participação no programa Tudo pela Audiência,	« 13 words »	esteve nos estúdios do Multishow para gravar participação no programa Tudo pela Audiência,
exibido		apresentado
por Tatá Werneck e Fábio Porchat.”	« 6 words »	por Tatá Werneck e Fábio Porchat. “
Acalento		Adoro
a Tatá e o Fábio. Me diverti	« 7 words »	a Tatá e o Fábio. Me diverti
bastante.		muito.
Encontrar com ela é sempre uma	« 6 words »	Encontrar com ela é sempre uma
<u>celebração”,</u> declarou Nicole.		<u>festa”,</u> disse Nicole.
A respeito da atração na TV a cabo, Nicole alegou: “Feliz demais em estar no elenco desse programa maravilhoso que é o Ferdinando Show. Considero o convite um presente de Deus e uma chance de aprender com quem tem experiência”. Prestes a estrear as gravações do programa Ferdinando Show nesta semana no canal Multishow, Nicole Bahls tirou a tarde de domingo para posar para as lentes do fotógrafo Moises Pazianoto, no Rio de Janeiro, onde posou para uma marca de roupas.		A gata, aliás, faz parte do novo casting da emissora. <u>Nicole estará</u> no programa Ferdinando Show.
Talita Werneck Arguelles, mais conhecida como Tatá Werneck é uma atriz, comedianta, apresentadora, repórter, baterista, tecladista e ex-VJ brasileira.		
A gata, aliás, faz parte do novo casting da emissora.		

other source
rewritten

WIKIPEDIA

DISLOCATION

How does our NLG architecture work?



Some metrics:

- ✓ Our injection of digital contents increased organic traffic at our customers upto 80% from the first month
- ✓ We have 35.000 unique visitors monthly at our own blog <http://entretenimientobit.com>
- ✓ Our blog has immediate indexation by Google, where our average position is 15 at Google Search Console
- ✓ Human readers and journalism agencies quote our posts

We are far ahead from our competitors in Spanish & Portuguese

English has been traditionally the target language for NLP frameworks. Thus the opportunity windows there were small and crowded.

Companies owning decent NLP solutions for English thought they could just translate their lexical resources into other languages. It doesn't work like that. Spanish & Portuguese have more challenging grammars than English and so specific grammar resources were needed.

State-of-art technology in NLP / NLU / NLG

Understanding Natural Language the way native speakers do