

Introducing RosaeNLG, an Open Source Natural Language Generation product

Virtual LF AI Day EU

June 22, 2020

NLG (Natural Language Generation)

Natural Language Generation (NLG) is transforming structured data into textual narratives, generally following predefined business rules.

- Mature technology used in production by large companies (Société Générale, BNPP, Moodys)
- Benefits: reliability, speed, production cost
- Automates the production of relatively repetitive texts, when based on structured data
- Requires:
 - input data
 - a NLG software
 - a **significant setup effort**: business rules, text templates

What are the Usecases for NLG

NLG automates the production of repetitive reports and texts based on structured data. For instance:

- describe a product and its features based on its characteristics (SEO)
- produce structured reports in the banking/financial industry e.g. risk reports, financial fund performance
- summarize situations e.g. client summary, preparation to meeting, meeting minutes
- generate personalized emails

NLG approaches

The template based approach is the most used today in production.

- NLG templates combine:
 - static texts
 - structures (e.g. conditions, loops, lists)
 - NLG functions (e.g. agreements, conjugations)
 - local processing using code (e.g. filtering, sorting)
- These templates are run by a specific template engine: the **NLG engine**
 - punctuation, capitalization
 - synonyms

Template example

Edit your RosaeNLG template: *based on "conjugation" en_US*

```
1 | he #[+verb(getAnonMS(), {verb: 'eat', tense: 'PAST'})]
2 itemz { separator: ', ', last_separator: 'and', end: '.' }
3   item
4     | a orange
5   item
6     | a pear
7   item
8     | #[+value('apple', { det: 'INDEFINITE', adj: {BEFORE: ['big', 'red'], SEP_BEFORE: '' } })]
9 | his favorite fruits are
10 - let fruits = ['avocado', 'mango', 'pear']
11 eachz fruit in fruits with { separator: ', ', last_separator: 'and also', end: '.' }
12   | #[+value(fruit, {number: 'P'})]
```

Render automatically

done!

Rendered texts: [show as html](#)

He ate an orange, a pear and a big red apple. His favorite fruits are avocados, mangoes and also pears.

RosaeNLG



Introduction

RosaeNLG is an open-source (MIT) NLG (Natural Language Generation) product.

Configuration uses templates which follow [Pug](#)'s popular syntax.

Local processing code in templates is written in JavaScript.

RosaeNLG is available in JavaScript or Java, or as a cloud-based SaaS.

RosaeNLG supports English, French, German, Italian and Spanish.

Packaging

Table 1. RosaeNLG is available under multiple packagings

Language	Library	API	Docker
JavaScript, TypeScript	<ul style="list-style-type: none">• npm package for node.js• JavaScript library (to generate texts directly in a browser)	<ul style="list-style-type: none">• node.js server• Lambda functions for AWS	yes
Java (JVM)	.jar published on Maven central	yes (.war)	yes

RosaeNLG Characteristics

RosaeNLG is based on [Pug](#), a fast, robust and widespread template engine.

Pug directly manages *if / else*, *case / when*, textual macros, etc.

RosaeNLG is **developer and IT friendly**:

- standard languages: Pug and JavaScript
- template development in VSCode with syntax highlighting
- benefits from the existing ecosystem (Git, JavaScript, node.js): easy to test, easy to CI/CD, easy to integrate

RosaeNLG offers **the same NLG features as the best vendor solutions** and some new features, such as the global optimization of synonyms: texts (or certain sections) can be generated several times and the best alternative is chosen.

Examples

Tire descriptions for Lizeo Group

- tire data and price monitoring worldwide specialist
- goal: generate unique original tire descriptions for > 10 000 tire references
- duplicate content challenge over different resellers

838 MT #1

Le 838 MT est destiné à des véhicules 4 roues motrices, il est développé pour un usage en tout-chemin.

Le 838 MT inséré par la marque Achilles est un pneumatique "été" prévu pour des véhicules de franchissement. Cette enveloppe est spécifiquement dédiée à être utilisée en dehors des sentiers battus.

Ce produit comporte des dimensions allant du 14 au 16 pouces. Il est disponible en 5 versions dimensionnelles. Ce pneu ne dispose pas de version spécifique pour un constructeur et est donc adapté à toutes les marques.

- Pneumatique au bon rapport performances/prix pour véhicules 4 roues motrices
- Économie mise en avant
- Robustesse en conditions off-road

random seed : 671

Dueler H/P Sport #1

Le pneumatique Dueler H/P Sport est un produit "été" du manufacturier Bridgestone. Ce pneumatique est prévu pour des véhicules de type SUV. Cette enveloppe est principalement dédiée à être utilisée en tout-chemin mais peut cependant s'adapter pour un usage routier.

Destiné à des véhicules de type SUV, le pneu "été" Dueler H/P Sport de la marque Bridgestone est spécifiquement développé pour une finalité sur routes et chemins cassants. Il peut malgré tout s'apprêter pour un usage routier.

Ce pneu bénéficie de diamètres qui vont du 16 au 21 pouces. Il possède dans sa gamme 123 dimensions. Ce pneu est disponible en versions spécifiques Porsche, Audi (AO), Mercedes (Mo), Maserati et BMW (*).

En cas de crevaison, la technologie "Roulage à plat" (autrement dit "Runflat") vous permettra de continuer à rouler de manière modérée sur une courte distance.

Ce pneumatique dispose de performances optimales pour une conduite familiale. Il bénéficie d'aptitudes importantes en termes de confort de conduite et de sécurité. C'est un produit fiable, ses capacités en matière de longévité sont excellentes.

- La Dueler H/P Sport haut de gamme appropriée aux SUV
- Prestations globales élevées pour une conduite familiale
- Freinage haute-performance sur sols secs et mouillés

random seed : 205

Database querying in natural language for Radicalbit

Radicalbit Natural Analytics (RNA) is a DataOps platform for Streaming Data Integration and Real-time Advanced Analytics.

- natural language query parsed with AWS
- queries NSDb (streaming oriented time-series database)
- query result is transformed into text using RosaeNLG

find me the number of tweets

The number of tweets is 1077. Do you want to filter this result?

hashtag is Amazon

Would you like to add another condition?

no thanks

When hashtag is Amazon, the number of tweets is 673.

find me the plan of listened songs

I've found 2 different values of plan, plan Paid occurred 12971 times, that is the 70 percent, while plan Free occurred 30 percent, with 5414 listened Songs, and in the last 24 hour instead, plan Paid was at 68 percent, with 3652 listened Songs and Free at 32 percent, with 1700.

Financial Fund performance

- comment the monthly performance of financial funds
- contributors and detractors per sector

Uni-Global - Equities Emerging Markets - AA-USD

Fund performance

The fund returned +3.0% (gross of fees, in USD terms) in July, strongly outperforming its benchmark by 80bp (gross of fees, in USD terms), which increased by 2.2%. From a country point of view, our stock selection was a positive contributor to relative performance while country allocation was a positive contributor to excess returns.

Largest contributors of the month

China (CHINA TELECOM CORP LTD-H and AGRICULTURAL BANK OF CHINA-H) and Brazil stocks selected for the portfolio added the most to the fund's performance.

In terms of absolute performance our positions in WALMART DE MEXICO SAB DE CV, INDIAN OIL CORP LTD and SOUTHERN COPPER CORP --- US were the standout gainers rising by 0.0%, 0.0% and 0.0%, respectively.

With a rise of 0.0%, ZHEN DING TECHNOLOGY HOLDING (Taiwan, Technology) was the top contributor to excess returns.

With a rise of 0.0%, WALMART DE MEXICO SAB DE CV (Food Retailing in Mexico) was also a solid contributor.

Finally, with a rise of 0.0%, INFOSYS LTD (India, Software) was also a solid contributor.

At sectorial level, our selection of Materials, Software and Banks sectors produced gains.

Develop using VSCode


The screenshot shows the Visual Studio Code interface with a file explorer on the left and a code editor in the center. The file explorer shows a project structure with folders like 'data', 'src', and 'templates', and files like 'index.js', 'nonreg.js', 'texts.js', and various '.pug' files. The code editor displays a Pug template file named 'ean.pug' with the following content:

```
1  mixin ean
2    if pneu.ean
3      synz
4        syn
5          | #[+subjectVerb(pneu, {verb: 'référer', tense: 'PASSE_COMPOSE', aux: 'ETRE', agree: pneu})]
6          | par le code EAN #[+value(pneu.ean)]
7
8        syn
9          | #[+subjectVerb(pneu, {verb: 'identifier', tense: 'PASSE_COMPOSE', aux: 'ETRE', agree: pneu})]
10         | par le code EAN #[+value(pneu.ean)]
11
12        syn
13          | la référence EAN de
14          | #[+value(pneu)]
15          | est #[+value(pneu.ean)]
16
17        syn
18          | sa référence EAN
19          | est #[+value(pneu.ean)]
```

Below the code editor, the terminal window shows the rendered HTML output:

```
1: node
d'autres en terme de consommation de carburant ! Ce pneu est de classe B, ce qui est à peine moins bien que la classe de référence A (l'échelle va de A - le mieux, à G - le moins bien).<p>Le niveau sonore de ce produit est considéré comme Moyen. Le bruit au roulement est de 72 dB, ce qui est bien moins que niveau maximum admissible qui est de 75 dB (pour des pneus de 205 mm). Sachez que comme il s'agit d'un pneu renforcé (XL), la limite légale est majorée de 1 dB.<p>L'indice de charge est de 94 : le poids maximum qu'un pneu donné pourra supporter est de 670 kg. À savoir : prenez impérativement un indice de charge <b>égal ou supérieur</b> à celui des pneus d'origine !<p>Cet article est référencé par le code EAN 4019238791648.<p>Vous devez choisir un indice de vitesse supérieur ou égal à celui d'origine. Cette marchandise vous permettra d'aller jusqu'à 210 km/h - ce qui est bien suffisant. En hiver, il est possible de monter un indice de vitesse inférieur (d'une lettre) à la monte de pneus été.<p>Ce pneu est toutes saisons : parfait si vous n'êtes pas contraint par un hiver très froid !<p>Il n'est pas "Run Flat", c'est-à-dire qu'on ne peut pas continuer à rouler en cas de crevaison. À savoir, un pneu non "Run Flat" est souvent moins bruyant !<p>Ce produit est garanti 1 an.</div>
```

Reference Documentation

RosaeNLG // Docs 

RosaeNLG / Reference Documentation / Value / Numbers

RosaeNLG

- What's New in RosaeNLG
- Quick Start
- Developer experience
- Tutorials
- Integration, node.js, Java, Docker etc.
- Reference Documentation
 - Value
 - Numbers
 - Dates
 - Nominal Groups
 - Simplified Syntax
 - Referring Expressions
 - Adjectives
 - Eachz and Itemz
 - Synonyms
 - Verbs
 - Possessives
 - hasSaid
 - Filter
 - Other and Helpers
 - Multilingual Templates
 - Node.js compatibility
 - About NLG and RosaeNLG
 - Advanced

Value on a Number

Using `value` on a number is the proper way to format and output a number.

You have various options for formatting:

- by default it will format the number accordingly to the locale: `562407` will output `562,407` in `en_US`, `562 407` in `fr_FR` (thanks to `numeral lib`)
- set `AS_IS` flag to `true` to avoid this formatting
- set `TEXTUAL` flag to `true` to transform the number into text: `#[+value(5500, {'TEXTUAL':true})]` will output *five thousand five hundred*
- set `ORDINAL_NUMBER` flag to `true` to transform the number into an ordinal number: `#[+value(21, {'ORDINAL_NUMBER':true})]` will output *21st*
- set `ORDINAL_TEXTUAL` flag to `true` to transform the number into an ordinal text: `#[+value(20, {'ORDINAL_TEXTUAL':true})]` will output *twentieth*
- use `FORMAT` to set a format directly used by `numeral`. See [numeral.js formats](#). This is very practical for currencies, %, etc.
- use `agree` for `ORDINAL_TEXTUAL` in `it_IT` and `es_ES`, for instance to have *prima* and not *primo* (default agreement is `M`)


Try it! (`en_US`)

```
1 P
2 | #[+value(562407)] /
3 | #[+value(5500, {'TEXTUAL':true})] /
4 | #[+value(21, {'ORDINAL_NUMBER':true})] /
5 | #[+value(20, {'ORDINAL_TEXTUAL':true})] /
6 | #[+value(104000, {'FORMAT': '0a$'})]
7
```

Test =>

<p>562,407 / five thousand five hundred / 21st / twentieth / 104k\$</p>

Compile and Render in a Browser

Home Save 

Export  Delete  Help 

Edit your RosaeNLG template:

```
15   battery: 3300
16 },
17 {
18   name: 'OnePlus 3T',
19   colors: ['Black', 'Gold', 'Gray'],
20   displaySize: 5.5,
21   screenRatio: 73.15,
22   battery: 3400
23 }
24 ];
25
26
27 mixin colors
28 | #[+value(phone, {'NOT_IT':true})]'s available #[+syn('colors', 'tints', 'tones')]
29 | are
30 eachz color in phone.colors with { separator:',', last_separator:'and', end:'.' }
31 | #{color}
32
33
34 mixin itHasADisplay
35 | #[+value(phone)] has a display with
36
37 mixin display
38 itemz { begin_with_general: 'itHasADisplay', separator:',', last_separator:'and', mix:true }
39 item
40 | a physical size of #[+value(phone.displaySize)] inches
41 item
42 | a screen-to-body ratio of #[+value(phone.screenRatio)] %
```

Render automatically

done!

Rendered texts ([show raw](#)):

OnePlus 5T: a great phone! The phone's available colors are Black, Red and White. This phone has a battery of 3,300 mAh. It has a display with a physical size of 6 inches and a screen-to-body ratio of 80.43 %.

OnePlus 5: a great phone! The phone has a display with a physical size of 5.5 inches and a screen-to-body ratio of 72.93 % along with a battery of 3,300 mAh. This phone's available tints are Gold and Gray.

I really love the new OnePlus 3T. The phone's available tones are Black, Gold and Gray. This phone has a display with a screen-to-body ratio of 73.15 % and a physical size of 5.5 inches along with a battery of 3,400 mAh.

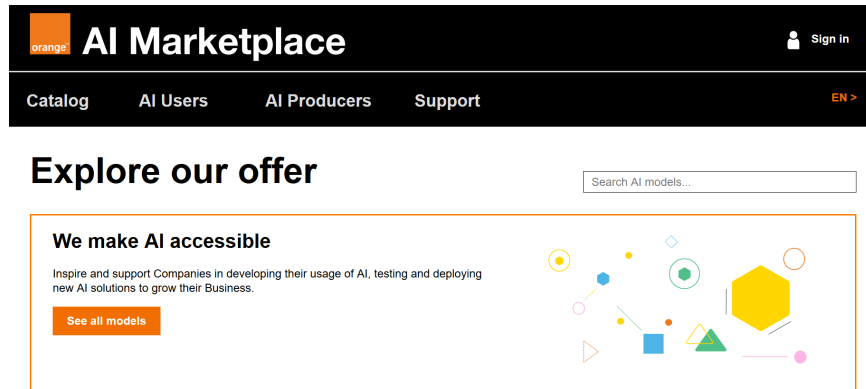
Support

- **RosaeNLG can be used without any restriction (MIT). All components are published in open-source.**
- Bugs and questions are reported directly to the author or via GitLab, and processed by the author in best effort.
- In practice, about 1 release per week (debug, update dependencies, security).
- Addventa (company specialized in NLG, based in Paris) provides commercial support on RosaeNLG:
 - support with SLA on RosaeNLG
 - Professional Services (consulting, development)
- The author can provide expertise on POC and small projects, directly or with Addventa.

RosaeNLG & The Orange AI Marketplace

RosaeNLG will be available (Q3 2020) on Orange AI marketplace:

- world class hosting by Orange
- self service interface for simple usecases (marketing, sales)



The screenshot shows the top navigation bar of the Orange AI Marketplace. It features the Orange logo and the text "AI Marketplace" on the left, and a "Sign in" button on the right. Below the navigation bar, there are links for "Catalog", "AI Users", "AI Producers", and "Support", along with a language selector "EN >". The main content area is titled "Explore our offer" and includes a search bar labeled "Search AI models...". Below the search bar, there is a section titled "We make AI accessible" with the text "Inspire and support Companies in developing their usage of AI, testing and deploying new AI solutions to grow their Business." and a "See all models" button. To the right of the text is a colorful graphic consisting of various geometric shapes (circles, squares, triangles, hexagons) connected by lines, representing a network or data flow.

<https://demo.ai-marketplace.orange-business.com/>

Detail

#	generated revenue	closed deals	outbound calls	client conversations
Henry C. Phillips	\$753 (\$759 => -1%)	78 (58 => +34%)	75 (31 => +142%)	85 (67 => +27%)
Ruth M. Peterson	\$475 (\$627 => -24%)	6 (22 => -73%)	32 (89 => -64%)	3 (27 => -89%)
Stella C. Mills	\$855 (\$34 => +2446%)	69 (73 => -5%)	89 (39 => +128%)	4 (80 => -95%)
Donald M. Nelson	\$737 (\$113 => +551%)	82 (71 => +15%)	88 (66 => +33%)	89 (55 => +62%)
Donna R. Howard	\$673 (\$200 => +237%)	92 (71 => +30%)	1 (16 => -94%)	91 (60 => +52%)
Tiffany E. Wilson	\$227 (\$202 => +12%)	24 (4 => +500%)	68 (63 => +8%)	8 (64 => -87%)
Geneva D. Bautista	\$275 (\$139 => +97%)	3 (43 => -93%)	45 (28 => +61%)	18 (97 => -81%)

Sales Overview

The total generated revenue this week is \$3,994, largely growing (+93%) compared to the week before (\$2,074 generated revenue). The best seller on generated revenue this week is Stella C. Mills with \$855 generated revenue, and this seller also has the best progression: +2446%. Congrats Stella C. Mills! On the other hand, the worst seller on generated revenue this week is Tiffany E. Wilson with \$227 generated revenue, while Ruth M. Peterson has the worst progression, from \$627 generated revenue to \$475.

The total closed deals this week is 354, almost stable (+4%) compared to the week before (342 closed deals). The best seller on closed deals this week is Donna R. Howard with 92 closed deals, while Tiffany E. Wilson has the best progression, from 4 closed deals to 24. On the other hand, the worst seller on closed deals this week is Geneva D. Bautista with 3 closed deals, and this seller also has the worst progression: -93%.

The total outbound calls this week is 398, growing (+20%) compared to the week before (332 outbound calls). The total client conversations this week is 298, largely shrinking (-34%) compared to the week before (450 client conversations).

Detail per Representative

Ranking

Donald M. Nelson is the best seller this week, with \$737 generated revenue and 82 closed deals, thanks to podiums on generated revenue, closed deals, outbound calls (88) and client conversations (89).

Contacts

Ludan Stoecklé

ludan.stoeckle@rosaenlg.org

Questions

Appendix

Contribute

Contribute on GitLab:

- create issues
- write documentation
- write code