

Discover how to include natural language generation in your applications

Professor Ehud Reiter, Chief Scientist, Arria NLG

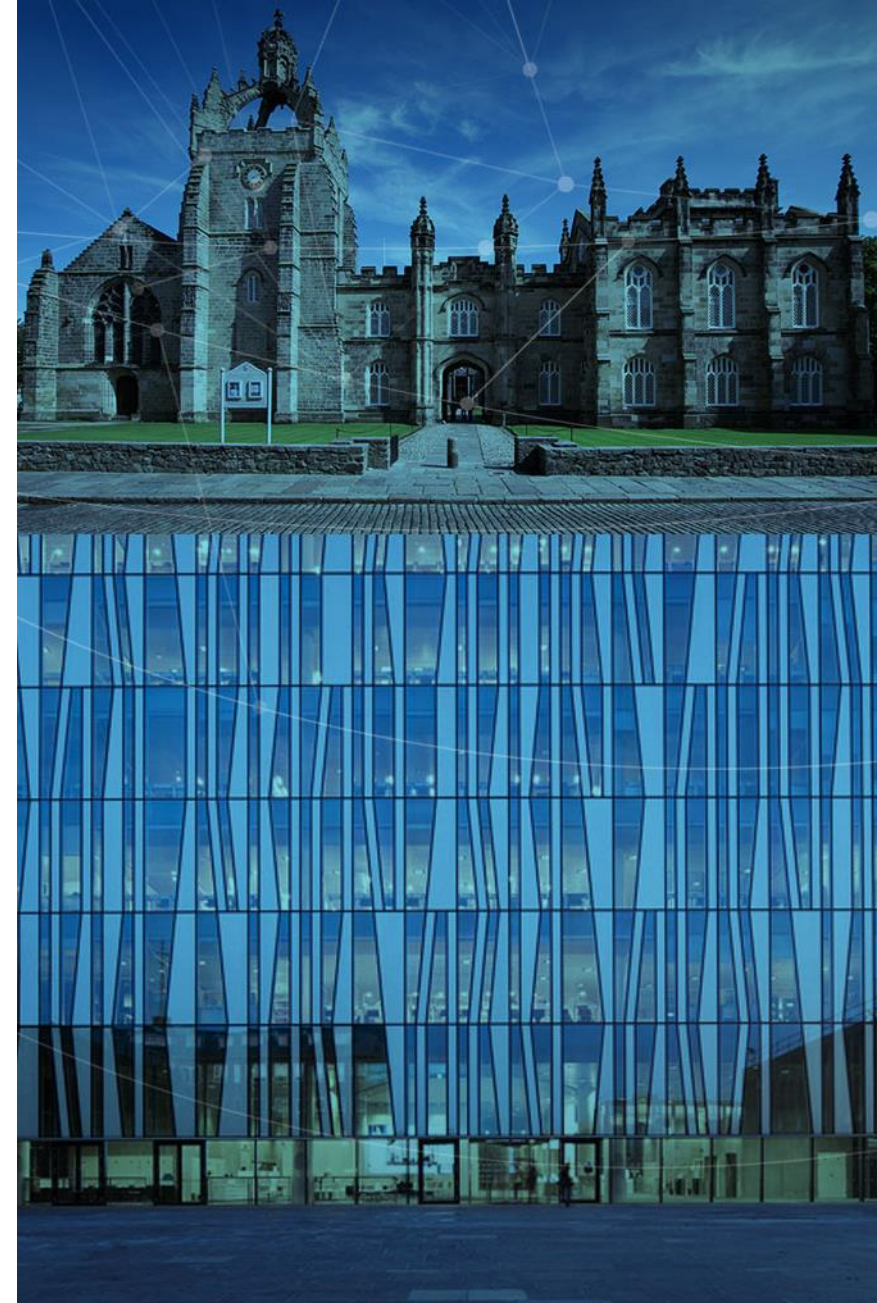
<https://github.com/arria-nlg/index-arria-including-nlg>

Welcome



ehudreiter.com
[@EhudReiter](https://twitter.com/EhudReiter)
arria.com

- Professor Ehud Reiter,
Chief Scientist of Arria NLG
- Professor of Computing Science at
University of Aberdeen
- Arria evolved from the spinout company
Data2Text which I co-founded in 2009
- Researching Natural Language Generation
for over 30 years
- Co-author of the text book “Building
Natural Language Generation Systems”



Discover How to Include NLG in Your Applications

- What is NLG
- NLG Methodology
- Arria NLG and its products
- Case study: Case Safety Reports (CSR)



What is Natural Language Generation?

- Software which can automatically produce reports, explanations, summaries, etc. of data, in well written and easy to read English (German, French, Spanish, Portuguese , ...)
- Artificial intelligence (AI) systems that can write!

Simple Example: Automatic Journalism

Expectant mums ignoring risk and lighting up while pregnant

ONE fifth of expectant mothers living in Dover, Deal and the Shepway district – which spans from Folkestone to the Romney Marsh – are smoking while pregnant.

There were 503 pregnancies reported in the second quarter of 2017, and 102 of these mothers (20.3 per cent) smoked throughout their pregnancy.

This is up from the 18.6 per cent of mothers who smoked in the same period of last year and from the 15.9 per cent of mothers who smoked four years ago.

The NHS wants to reduce the number of mothers who smoke while pregnant to six per cent – but the rate in the south Kent coast NHS commissioning area, which effectively covers most of east Kent, has increased even as numbers have gone down nationally.

The Department of Health has been targeting pregnant smokers due to the risk of serious health complications for both mother and child.

When a woman smokes during pregnancy it restricts the flow of oxygen to the unborn child.



Dover Express newspaper:

One fifth of expectant mothers living in Dover, Deal, and the Shepway district – which spans from Folkestone to the Romney Marsh – are smoking while pregnant,

There were 503 pregnancies reported in the second quarter of 2017, and 102 of them (20.3 per cent) smoked throughout their pregnancy...

Moderate Example: Weather Forecast

London Heathrow Airport

[Change table layout](#)

Tue 4 Mar

Wed 5 Mar

Thu 6 Mar

Fri 7 Mar

Sat 8 Mar

06:00 Wed 05 Mar 2014 - 06:00 Thu 06 Mar 2014

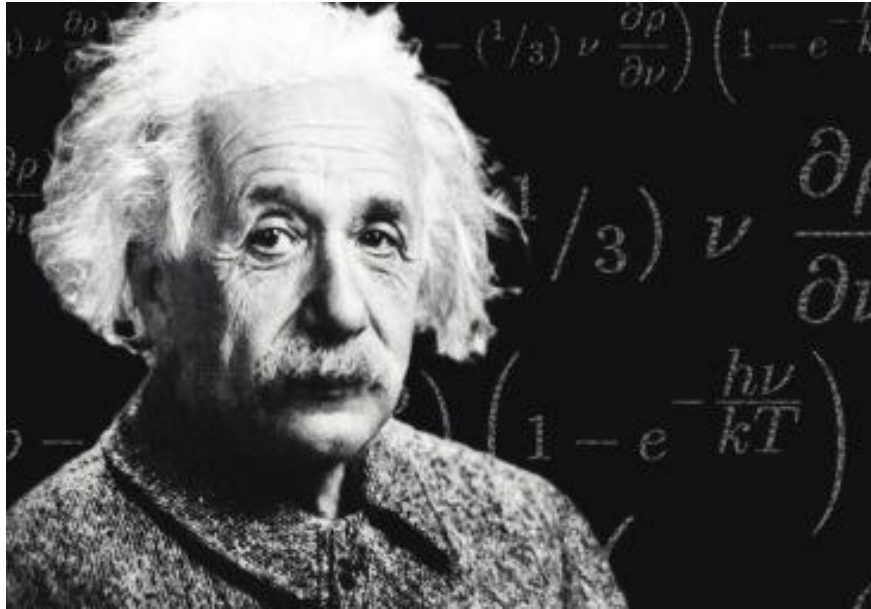
Sunshine from mid-morning and into the afternoon. Staying dry, but becoming cloudier from early evening and into Thursday. It is likely to feel milder than on Tuesday with a maximum temperature during the afternoon in the region of 11C and a minimum temperature overnight of around 6C. Light winds throughout.

UK local time	Warnings for Greater London	Weather	Precip. (%)	Temp (°C)	Feels like (°C)	Wind speed & direction (mph)	Wind gusts (mph)	Visibility	Humidity (%)	UK local time	Daily air quality index (BETA)
0000	No warnings		<5	4	3	4	No gusts	Moderate	90	0	
0300	No warnings		<5	3	2	4	No gusts	Moderate	92	0	

Complex Example: Market Stress Test (with IBM)

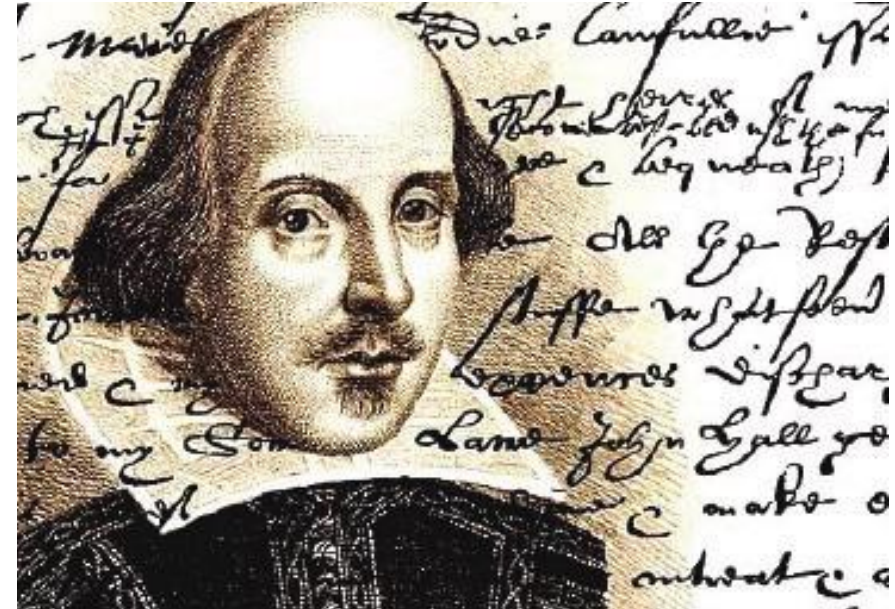
Example Data	Example Narrative
<pre>1 { 2 "scenario": { 3 "inputRiskFactorId": "CX_EQI_SPDJ_USA500_BMK_U 4 "inputRiskFactorName": "S&P 500 Index", 5 "inputMagnitudePercentage": "0.9" 6 }, 7 "correlations": [8 { 9 "category": "Index", 10 "correlationValue": 0.47733963064670165, 11 "all": [12 { 13 "correlationValue": 0.38411312889752697, 14 "factorId": "CX_EQI_AUSX_Sydney_BMK_AUD 15 "factorName": "ASX All Ordinaries Index" 16 "regionName": "APAC" 17 }, 18 { 19 "correlationValue": 0.5918764909782132,</pre>	<h2>Overview</h2> <p>Driven by your investments in Frontier Communications Corp and the Technology Sector, your portfolio could decrease 0.01% given the S&P 500 Index were to dive 10%. However, gains are expected in Anheuser-Busch Inbev Finance Inc. and Fedex Corp.</p> <h2>Detailed Analysis</h2> <p>Your greatest issuer loss came from Frontier Communications Corp which fell by 0.24% (339% of the total change in the portfolio). The next biggest was International Business Machines Co (0.11% decrease). Your largest rising issuer came from Fedex Corp in your</p>
Request Type: application/json	Response Type: application/json

Einstein and Shakespeare in concert



IBM

Watson is a cognitive technology that can think like a human.

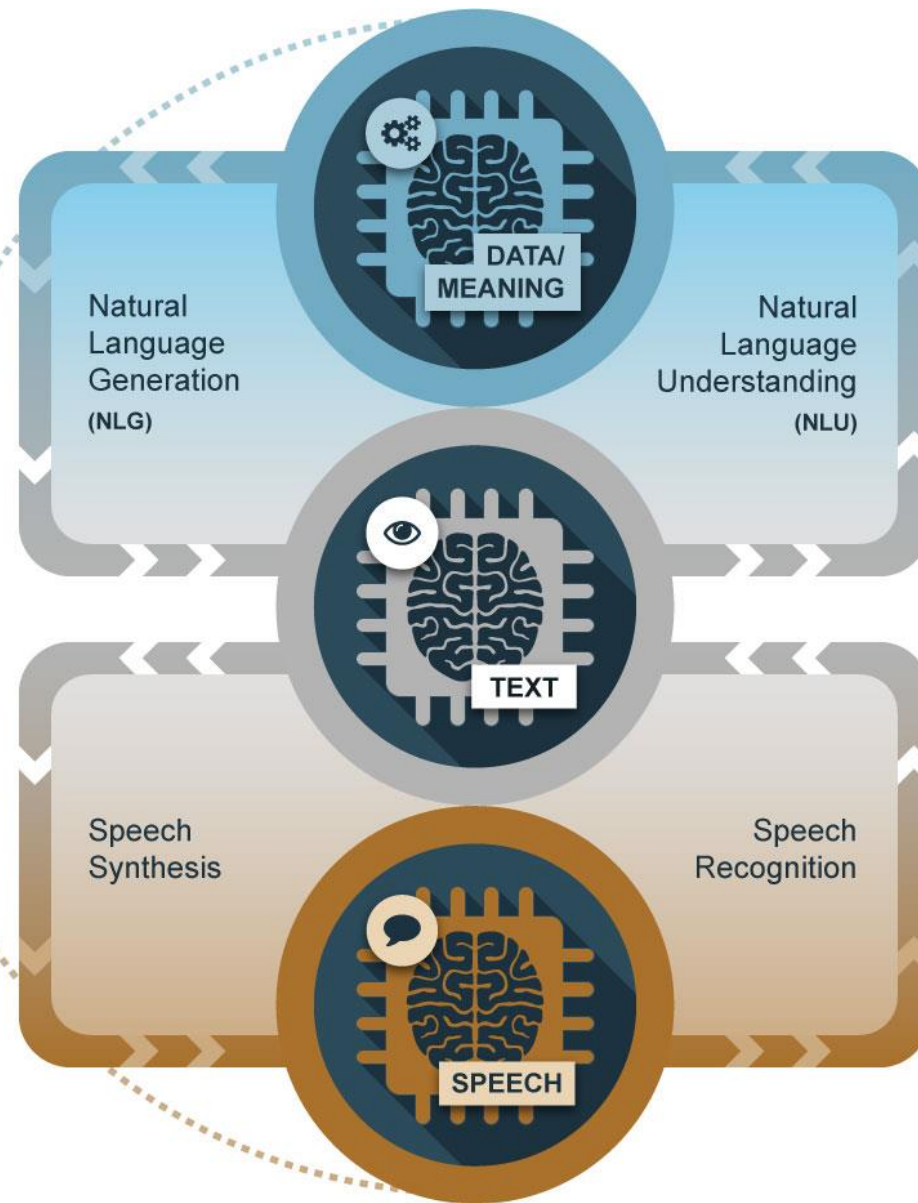


ARRIA

NLG is a cognitive technology that can communicate like a human.

NLP, NLU and NLG

NLP



Exciting uses of NLG

- Explaining data to people
 - So people feel empowered instead of drowned by “big data”
 - (My) research shows texts summaries can be more effective than visualisations
- Automated reporting
 - (My) research shows users sometimes prefer computer NLG texts over manually written texts, because NLG texts are more consistent and less idiosyncratic
- Natural dialogue in chatbots, dialogue systems, computer games, etc
- Assistive technology – helping people who cannot see and/or talk

Discover How to Include NLG in Your Applications

- What is NLG
- NLG Methodology
- Arria NLG and its products
- Case study: Case Safety Reports (CSR)



Arria Methodology for Building NLG Systems

- Covers requirements, design, coding, testing.
- Used in-house at Arria
- We are developing training videos and courses for partners and clients

Methodology: Requirements

- Usual techniques for requirements analysis
 - *Define use cases*: often use structured workshops, focusing on pain points in reporting or decision making
 - *Rapid prototyping*: especially important since many clients do not have a good idea of what NLG can and cannot do.
- Also use *corpus analysis*, which is more specific to NLG

Corpus Analysis

- Ask client for *target texts* that the application should produce
- Use *corpus analysis* to link phrases and sentences to input data
- Identify segments which are difficult/impossible to generate from the data
- **How is my portfolio doing today?**
Hello, Jane. So far today, your brokerage accounts portfolio is up .3%, and your highest performing account is your Roth Contributory IRA with a .47% increase. Your largest increase has been with Microsoft (MSFT), which has increased by 2.64%. In your Roth Contributory IRA account, most of your stocks have increased in value, but a few had a down day, including Apple (APPL) which is down .34%.

Methodology: Design

- Most NLG systems include *analytics* and *linguistic* processing
 - Analytics: come up with insights (*Einstein*)
 - Linguistic: express information well in language (*Shakespeare*)
- Key design decision is interface between the two
- Can be driven by analysis of examples in target texts

portfolioPerformance message

- changeDirection (UP or DOWN)
- percentageValueChange (float)
- Analytics (Einstein): compute message from stock data
- Linguistic (Shakespeare): produce “So far today, your brokerage accounts portfolio is up .3%,”

Methodology: Test

- Testing NLG systems most similar to testing other types of software
- One challenge specific to NLG is variation
 - NLG systems often vary narratives to avoid repetition
 - Your account rose 0.3% vs Your account went up 0.3%
 - Testing needs to look at all variants

Discover How to Include NLG in Your Applications

- What is NLG
- NLG Methodology
- Arria NLG and its products
- Case study: Case Safety Reports (CSR)



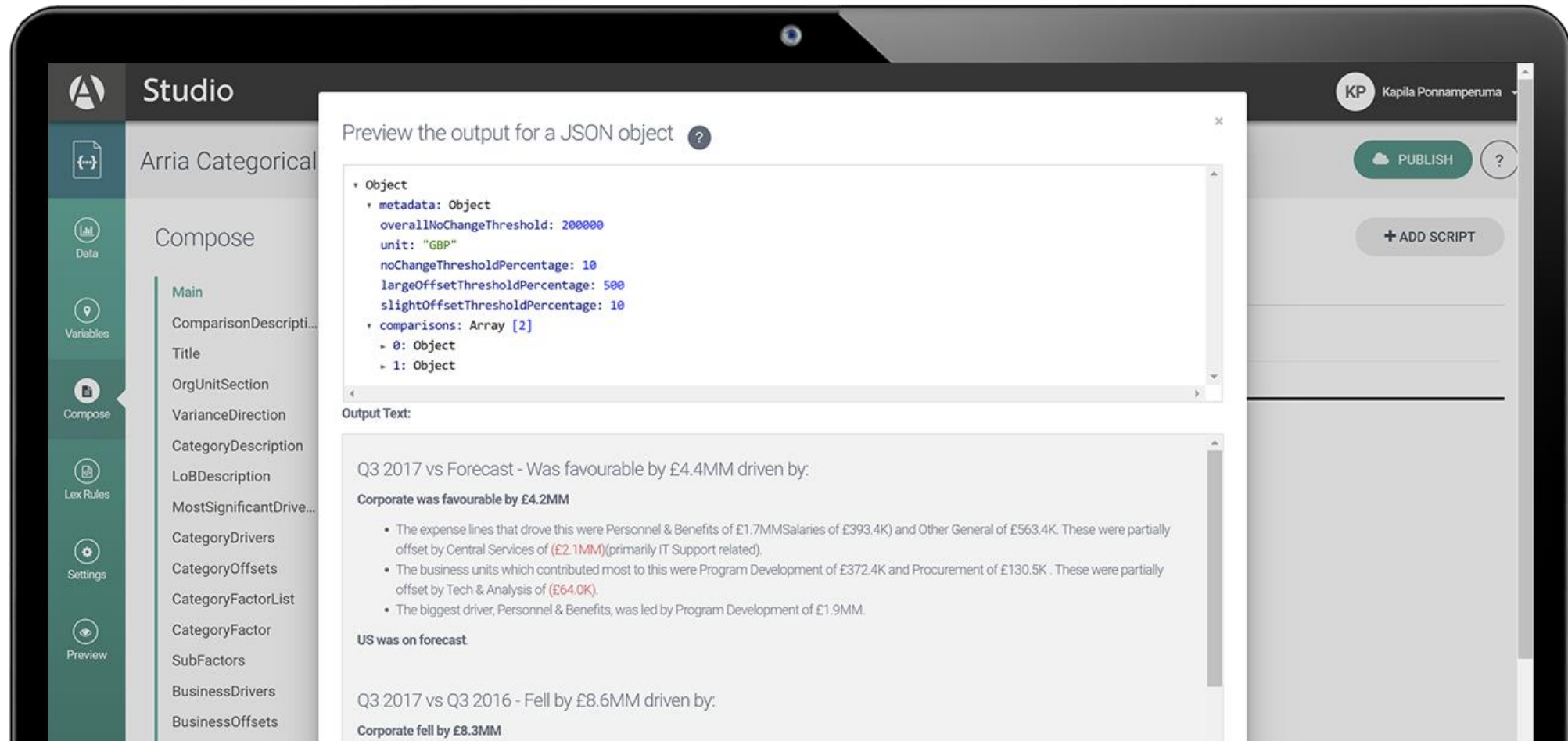
About Arria NLG

- Mission: leader in NLG and real-time data storytelling
- Founded by myself and others from the University of Aberdeen
- 13 patents and continued innovation leading to new ones
- Global Presence with offices in New York, London, Aberdeen, Sydney and Auckland

Arria Developer Tools

- NLG Studio: web-based system for building NLG apps
- NLG SDK: Java SDK for building NLG apps
- NLG Microservices: web-based services for simple generic NLG tasks
- NLG Narrative APIs: web-based NLG apps for specific use cases
- NLG Videos and Training Material

Arria NLG Studio



Arria NLG SDK

The screenshot displays the Arria NLG SDK website. On the left is a dark sidebar with the 'A SDK' logo and a list of navigation items: '6.3.0', 'Quick Start Guide', 'Release Notes', 'Tutorials', 'Quick Reference Sheets', 'Reference Manuals', 'Microservices', 'NLG Authoring Tools', 'Other Guides', and 'API'. The main content area has a light blue header with a search icon and a menu icon. Below the header is a large banner for the 'Quick Start Guide' with the text 'Get started on your NLG SDK journey by completing this Quick Start Guide. It teaches you how to install the software and build your first system.' and a 'Get Started >' button. Underneath is a section for 'Latest Release Notes' for version 6.3.0, describing the 'Lex Rule Generator (LRG)' and including a 'Read More' button and a 'Lex Rule Generator' button. At the bottom, there are three small preview images: one showing XML-like code, one showing a futuristic interface, and one showing a geometric pattern.

A SDK

6.3.0

- Quick Start Guide
- Release Notes
- Tutorials
- Quick Reference Sheets
- Reference Manuals
- Microservices
- NLG Authoring Tools
- Other Guides
- API

Quick Start Guide

Get started on your NLG SDK journey by completing this Quick Start Guide. It teaches you how to install the software and build your first system.

[Get Started >](#)

Latest Release Notes

Version 6.3.0 introduced the new [Lex Rule Generator \(LRG\)](#), – a web app that takes a sentence and will produce a Lex Rule for you. Find out about the LRG and a whole host of other new features by reading the release notes.

[Read More](#)

[Lex Rule Generator](#)

```
<Plugins>
<Reference>
<Instances>
  <NamedEntity type="configuredNamedE
  <ShortNamesAllowed value="true"
```

Arria NLG Microservices

ARRIA NLG STORE

Log out

Narrative APIsMicroservicesNLG ToolsMy Account

Get Started

aOrAn

countable

currency

genderedPronoun

list

pluralNoun

pluralVerb

pronoun

sentence

spelledNumber

verbGroup

aOrAn

The aOrAn service determines whether to use "a" or "an" before a word, e.g. aOrAn with "umbrella" produces "an", aOrAn with "ukulele" produces "a".

Parameters	Required/Optional	Description
word	Required	The word for which the use of "a" or "an" is to be determined.
apikey	Required	Your API key.

Examples

1. The end point returns a for the word *university*.
Request URL: `https://micro.arria.com/services/rest/english/aOrAn?word=university&apikey=<API-KEY>`
Response: a [Try it](#)

2. The end point returns an error message if the word parameter is empty:
Request URL: `https://micro.arria.com/services/rest/english/aOrAn?word=&apikey=<API-KEY>`
Response: `IllegalArgumentException : The word parameter for aOrAn is invalid.` [Try it](#)

cURL

```
curl "https://micro.arria.com/services/rest/english/aOrAn?word=university&apikey=<API-KEY>"
```

Reset ↺

Example Input ?
`word=university`

Try it

Example Output ?
`a`

Arria NLG Narrative APIs

The screenshot displays the Arria NLG Narrative API interface. At the top, there are tabs for 'Portfolio Management', 'Financial Services', and 'Life Sciences'. Below these, a modal window titled 'Predictive Market Stress Testing' is open. It contains a description: 'Pairs with IBM analytics to describe the effect of a stress test on your investment portfolio.' and a link to 'See documentation'. Below this, it says 'Your Predictive Market Stress Testing Formatted URL. Copy and paste this into your application or service.' and shows a POST request to 'https://stresstesting-narrativeapi.arria.com/services/rest/fullnarrative'. The modal is divided into two main sections: 'Example Data' and 'Example Narrative'. The 'Example Data' section shows a JSON object with a scenario, input risk factors, and correlations. The 'Example Narrative' section shows an overview and a detailed analysis of the stress test results. At the bottom of the modal, there is a 'Generate Narrative Preview' button.

Predictive Market Stress Testing

Pairs with IBM analytics to describe the effect of a stress test on your investment portfolio.
[See documentation](#)

Your Predictive Market Stress Testing Formatted URL. Copy and paste this into your application or service.

POST `https://stresstesting-narrativeapi.arria.com/services/rest/fullnarrative`

Example Data

```
1 - {
2 -   "scenario": {
3 -     "inputRiskFactorId": "CX_EQI_SPDJ_USA500_BMK_U",
4 -     "inputRiskFactorName": "S&P 500 Index",
5 -     "inputMagnitudePercentage": "0.9"
6 -   },
7 -   "correlations": [
8 -     {
9 -       "category": "Index",
10 -      "correlationValue": 0.47733963064670165,
11 -      "all": [
12 -        {
13 -          "correlationValue": 0.38411312889752697,
14 -          "factorId": "CX_EQI_AUSX_Sydney_BMK_AUD",
15 -          "factorName": "ASX All Ordinaries Index",
16 -          "regionName": "APAC"
17 -        },
18 -        {
19 -          "correlationValue": 0.5918764909782132,
20 -          "factorId": "CX_EQI_USA500_BMK_US",
21 -          "factorName": "S&P 500 Index",
22 -          "regionName": "NA"
23 -        }
24 -      ]
25 -     }
26 -   ]
27 - }
```

Request Type: application/json

Example Narrative

Overview

If there were a 10% dive in the S&P 500 Index your portfolio might fall 0.01%, amplified by your investments in Frontier Communications Corp and the Technology Sector. However, increases are expected in Anheuser-Busch Inbev Finance Inc. and Fedex Corp.

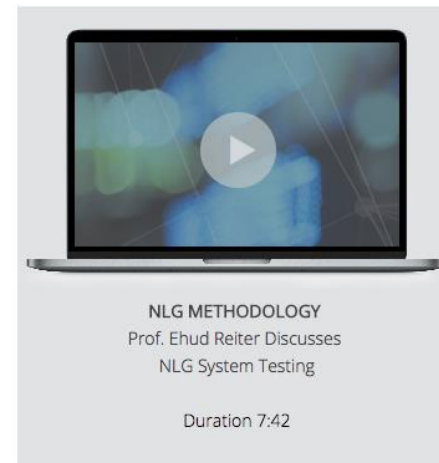
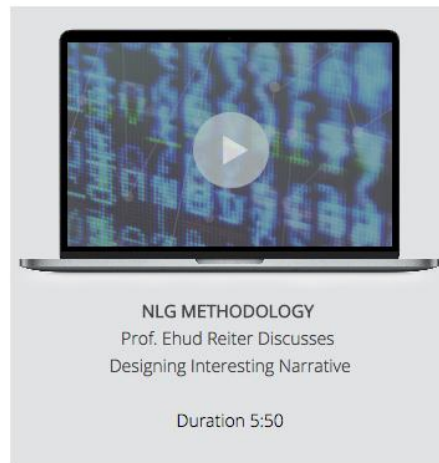
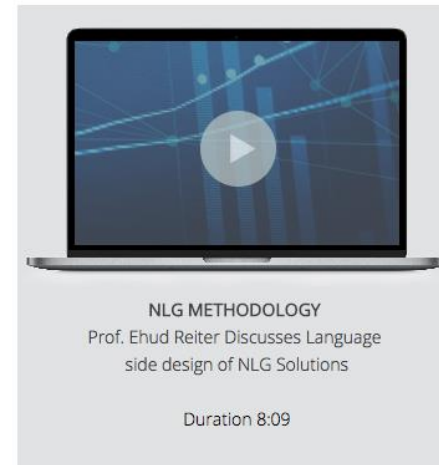
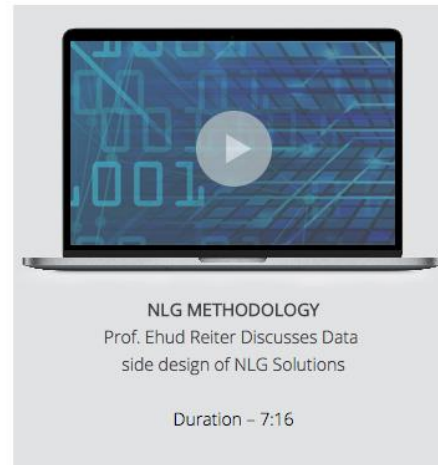
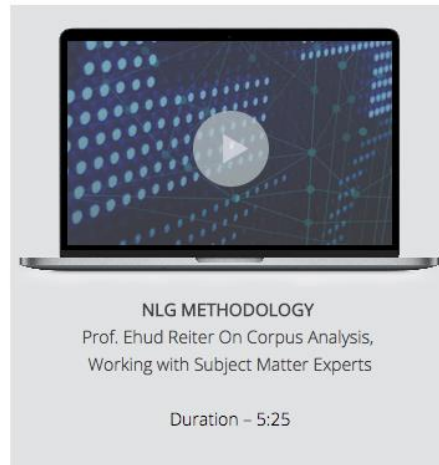
Detailed Analysis

Your largest issuer loss came from Frontier Communications Corp which fell by 0.24% (339% of the total change in the portfolio). The next greatest was International Business Machines Co (0.11% decrease). Your biggest rising issuer came from Fedex Corp in your

Response Type: application/json

[Generate Narrative Preview](#)

Arria NLG Videos and Training Material



Discover How to Include NLG in Your Applications

- What is NLG
- NLG Methodology
- Arria NLG and its products
- Case study: Case Safety Reports (CSR)



Case Study: Case Safety Reports (CSR)

- Pharmaceutical companies need to file *case safety reports* (CSR) with regulators when there is an adverse reaction to a medication
 - Part of *pharmacovigilance*
 - Both during clinical trials and after medication is approved
- Arria is working with a number of clients on partially automating the production of CSR
 - Assuming relevant information has been gathered
 - Arria's focus is on producing an acceptable report for regulators
 - Reporting is faster, cheaper, more accurate, and more consistent
 - NLG automation means clients can produce interim as well as final reports

CSR in Studio

- Open project
- Look at data
- Look at preview output, link to data
- Look at variables
- Look at ATL
- Modify first para, including pronouns

Try It Yourself with a Free Trial!

- Studio: <http://studio.arria.com/>
- Microservices and Narrative services: <https://nlgapi.arria.com/>
- Videos: <https://www.arria.com/space/video/>

Arria at Index

- Booth in exhibition
- Theatre talk tomorrow (Thurs) at 1205
 - Talk to Your Users with Natural Language Generation



ARRIA

NATURAL LANGUAGE GENERATION

NORTH AMERICA | EMEA | APAC