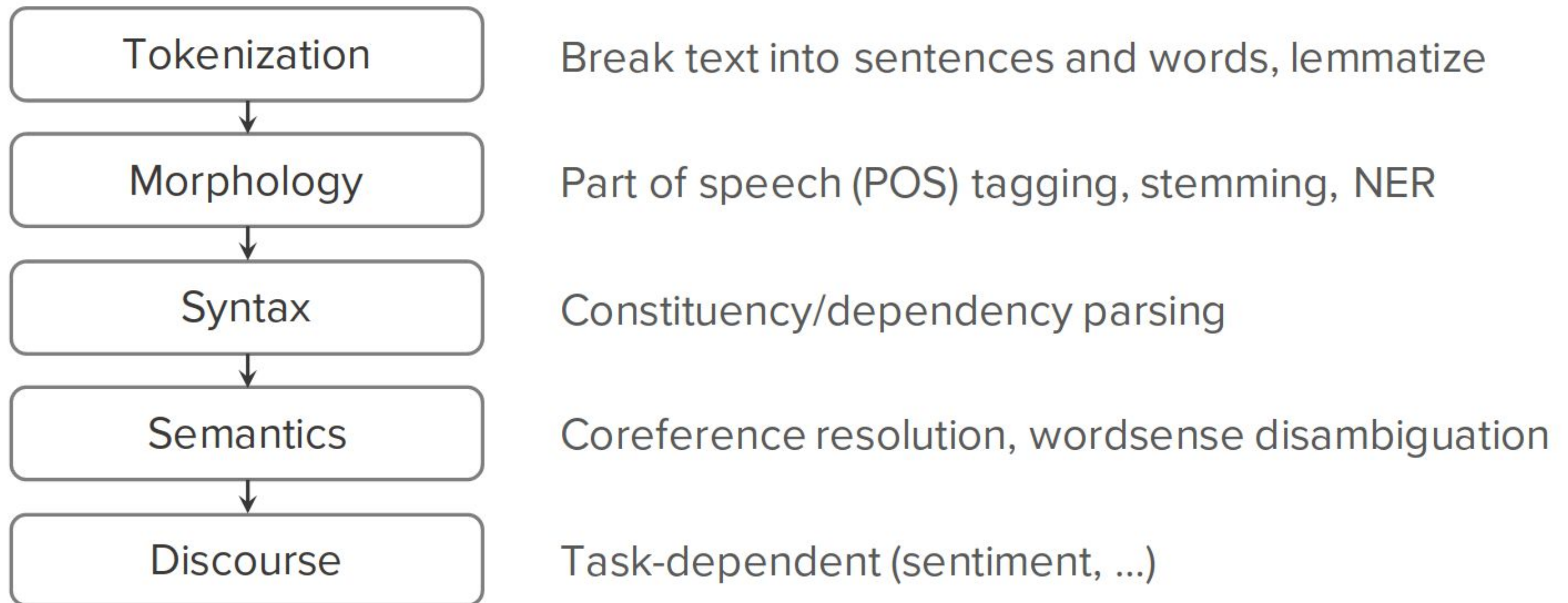



Análise de Mídias Sociais e Mineração de Texto

Aquisição e Modelagem de Dados

Laura de Oliveira F. Moraes






DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT


The home of the U.S. Government's open data


Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.


Credit Card Complaints

BROWSE TOPICS

Agriculture

Climate

Health

Local Government

BRASIL

Serviços

Participe

Acesso à Informação

Legislação

Canais

Ministério da Transparência e Controladoria-Geral da União

Portal da Transparência

GOVERNO FEDERAL

Perguntas frequentes

Contato

Glossário

Links

Manual de navegação

CONSULTAS

Despesas

Receitas

Convênios

Empresas Sancionadas

Empresas Punidas

Entidades Impedidas

Servidores

Imóveis Funcionais

Beneficiários L10.559/02

GRÁFICOS E DOWNLOADS

Download de Dados

Portal em Gráficos

INFORMAÇÕES

Sobre o Portal

Dados do Portal

Aprenda Mais

Notícias

Controle Social

Licitações e Contratos

DESPESAS

RECEITAS

CONVÊNIOS

SANÇÕES

SERVIDORES

+ CONSULTAS

Despesas - Empenhamento

Despesas - Pagamento

Transferências de Recursos

Consultas Temáticas

DADOS DO PORTAL

sign in

become a supporter

subscribe

search

jobs

dating

more

International edition

the guardian

browse all sections

home

UK

world

sport

football

opinion

culture

business

lifestyle

fashion

environment

tech

travel

Open journalism Datablog

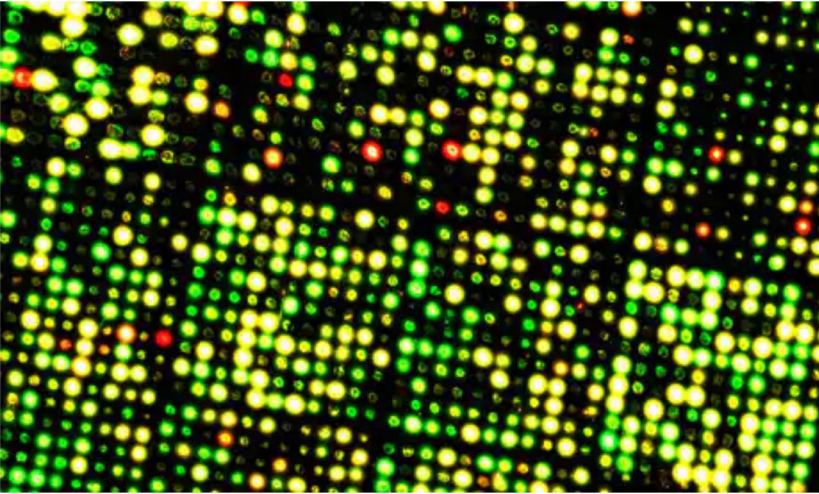
Simon Rogers

Thursday 20 September 2012 08:00 BST

Open data journalism

Why does data journalism need to be open journalism too? **Simon Rogers** explains how we do it on the Datablog

More data journalism and data visualisations from the Guardian



Can charities and non-governmental organisations use their information more effectively? Photograph: Deco / Alamy/Alamy

Data is everywhere: from governments publishing billions of bytes of the stuff, to visual artists creating new concepts of the world through to companies building businesses on the back of it.

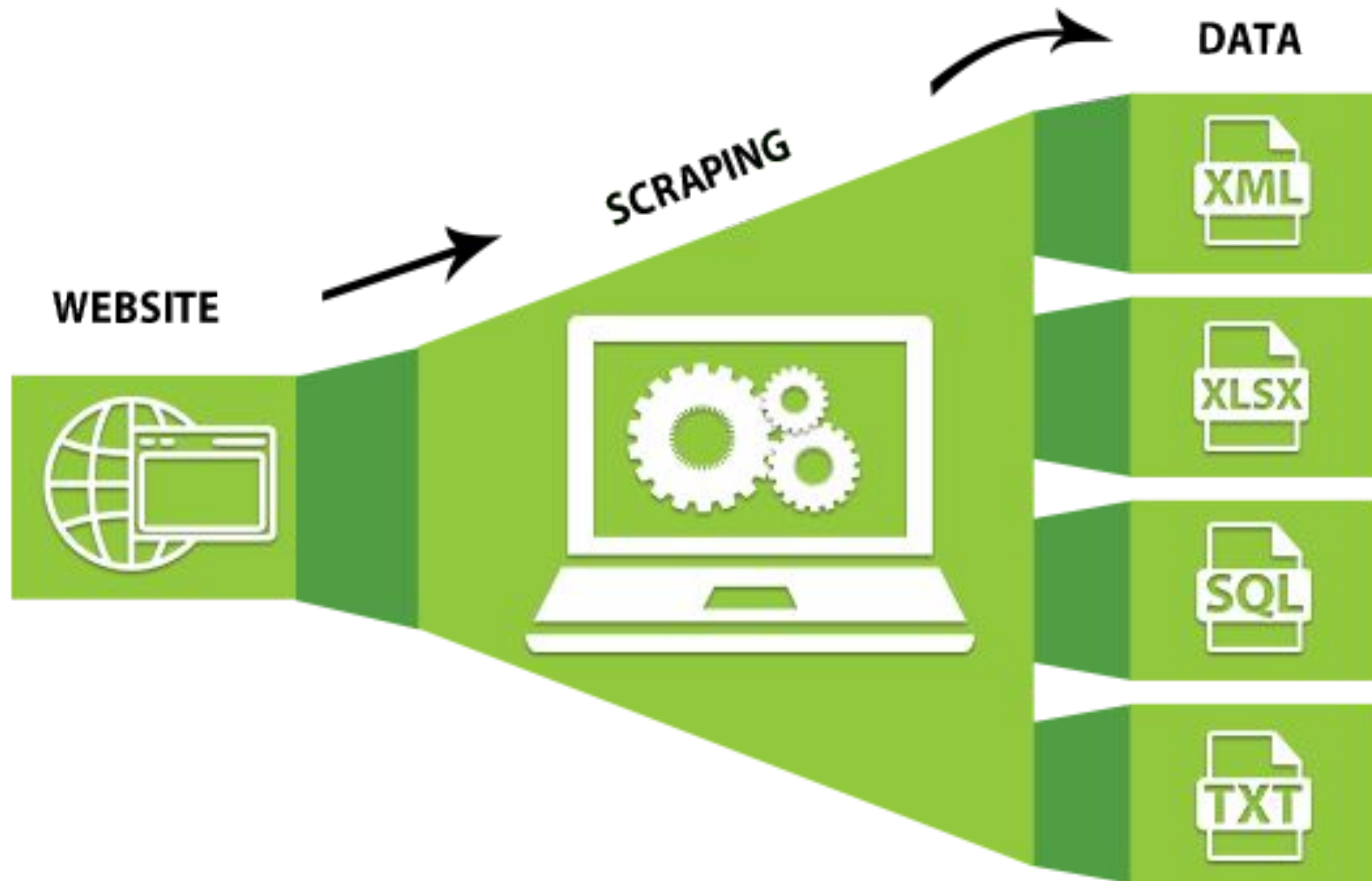
And governments want to be data journalists too – the business for others has never

Advertisement

Pare de seguir leads que não produzem resultados

Permita que o PipeDrive lhe mostre o caminho

Teste gratuitamente



- **Copiando e colando:** devagar e ineficiente. Envolve humanos analisando e copiando os dados
- Manualmente identificando padrões no texto: através de **regular expressions**.
- **APIs:** muitos sites como Facebook, Twitter, LinkedIn disponibilizam APIs públicas ou pagas que podem ser acessadas através de chamadas http em um formato específico.
- Parseamento de **DOM:** através de navegadores web, é possível recuperar dinamicamente o conteúdo gerado pelos scripts no cliente. Gera-se uma árvore e é possível navegar através dela e recuperar as partes de interesse em uma página.

- **Copiando e colando:** devagar e ineficiente. Involve humanos analisando e copiando os dados
- Manualmente identificando padrões no texto: através de **regular expressions**.
- **APIs:** muitos sites como Facebook, Twitter, LinkedIn disponibilizam APIs publicas ou pagas que podem ser acessadas através de chamadas http em um formato específico.
- Parseamento de **DOM:** através de navegadores web, é possível recuperar dinamicamente o conteúdo gerado pelos scripts no cliente. Gera-se uma árvore e é possível navegar através dela e recuperar as partes de interesse em uma página.

<https://developer.twitter.com/>


[Developer](#)
[Use cases](#)
[Products](#)
[Docs](#)
[More](#)
[Labs](#)

[Apply](#)

[Sign In](#)

Developers

Tap into what's happening.

Publish and analyze Tweets, optimize ads, and create unique customer experiences.



Announcing Twitter Developer Labs: [Learn more >](#)

Use cases

Advertise

Programmatically create and manage Twitter ad campaigns.

[Learn more >](#)

Publish and curate

Tell great stories with Twitter content.

[Learn more >](#)

Analyze

Evaluate Twitter data to inform business decisions.

[Learn more >](#)

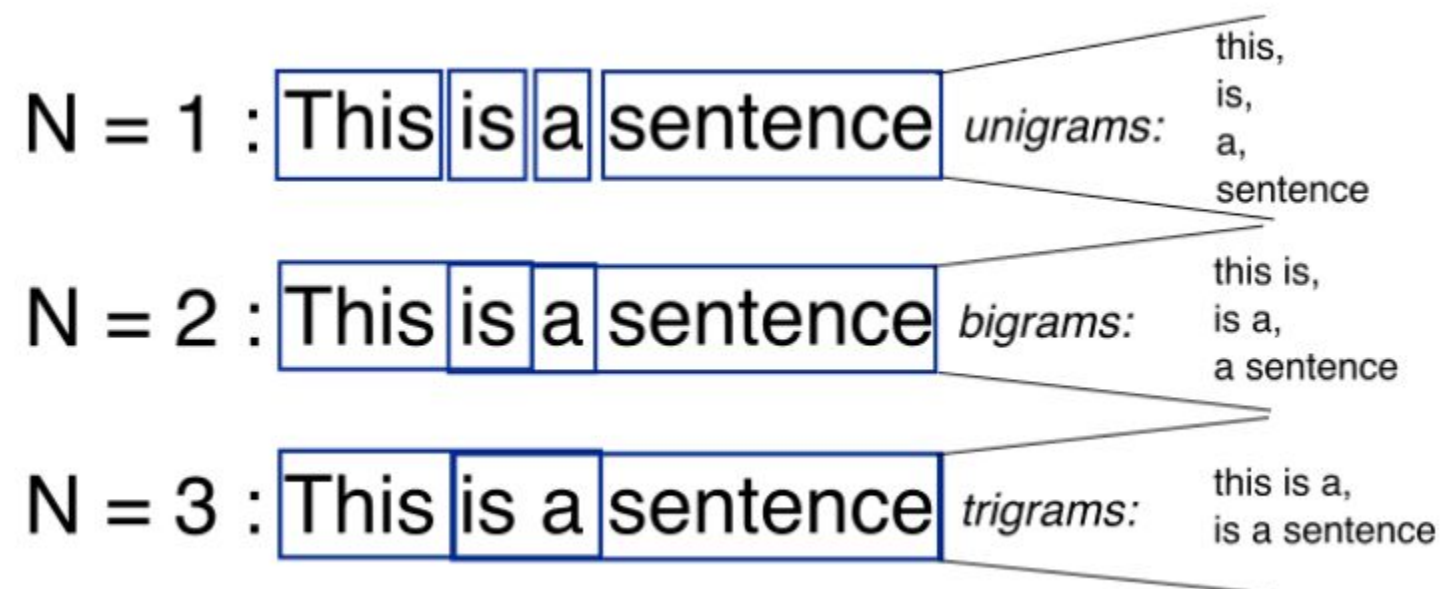
Engage

Create connections with conversational experiences.

[Learn more >](#)

the dog is on the table

0	0	1	1	0	1	1	1
are	cat	dog	is	now	on	table	the




```
def light(switchA, switchB):
    if switchA == 1 and switchB == 1:
        return True
    else:
        return False
```

```
def lessthan5(num_list):
    new_list = []
    for item in num_list:
        if item < 5:
            new_list.append(item)
    return new_list
```

```
def light switchA switchB is_block
    is_indent if switchA is_op_logic is_number and switchB is_op_logic
is_number is_block
        is_indent return True
    is_dedent else is_block
        is_indent return False
    is_dedent
is_dedent
```

```
def lessthan5 num_list is_block
    is_indent new_list is_attribution is_list
    for item in num_list is_block
        is_indent if item is_op_logic is_number is_block
            is_indent new_list is_class append item
        is_dedent
    is_dedent return new_list is_dedent
```

N-grams → Bag of words

term\document	doc #1	doc #2
is_number	2	1
is_op_logic	2	1
if	1	1
append	0	1
is_attribution	0	1
for	0	1
is_list	0	1
is_op_logic is_number	2	1
is_attribution is_list	0	1
is_attribution is_list for	0	1