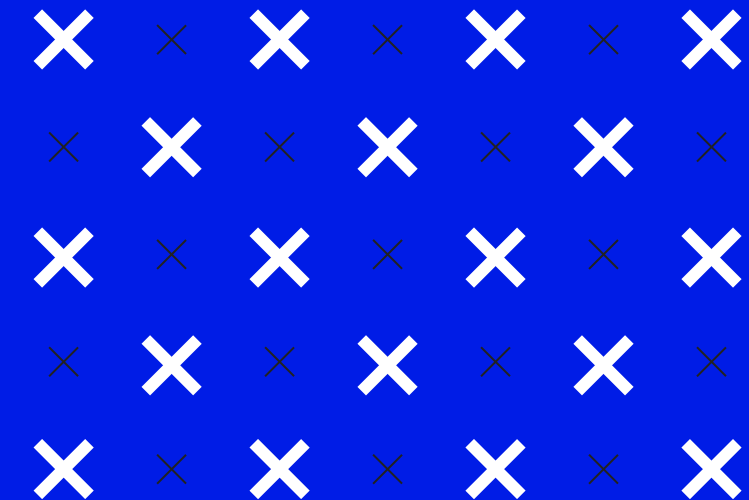


## MÓDULO 04

# HTML Markup



**mentorama.**

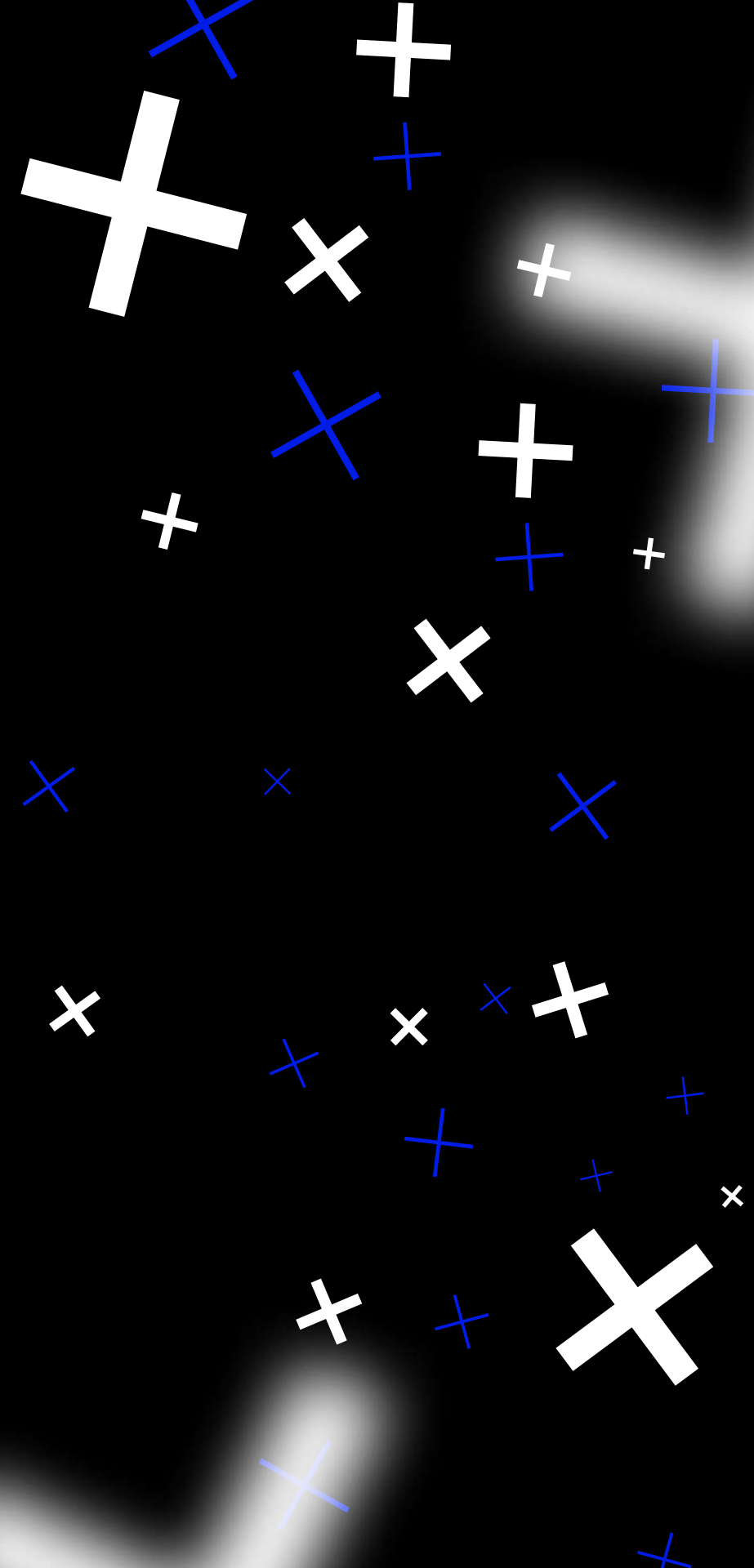
**mentorama.**

# Apresentação

HTML Markup

**mentorama.**

mentorama.

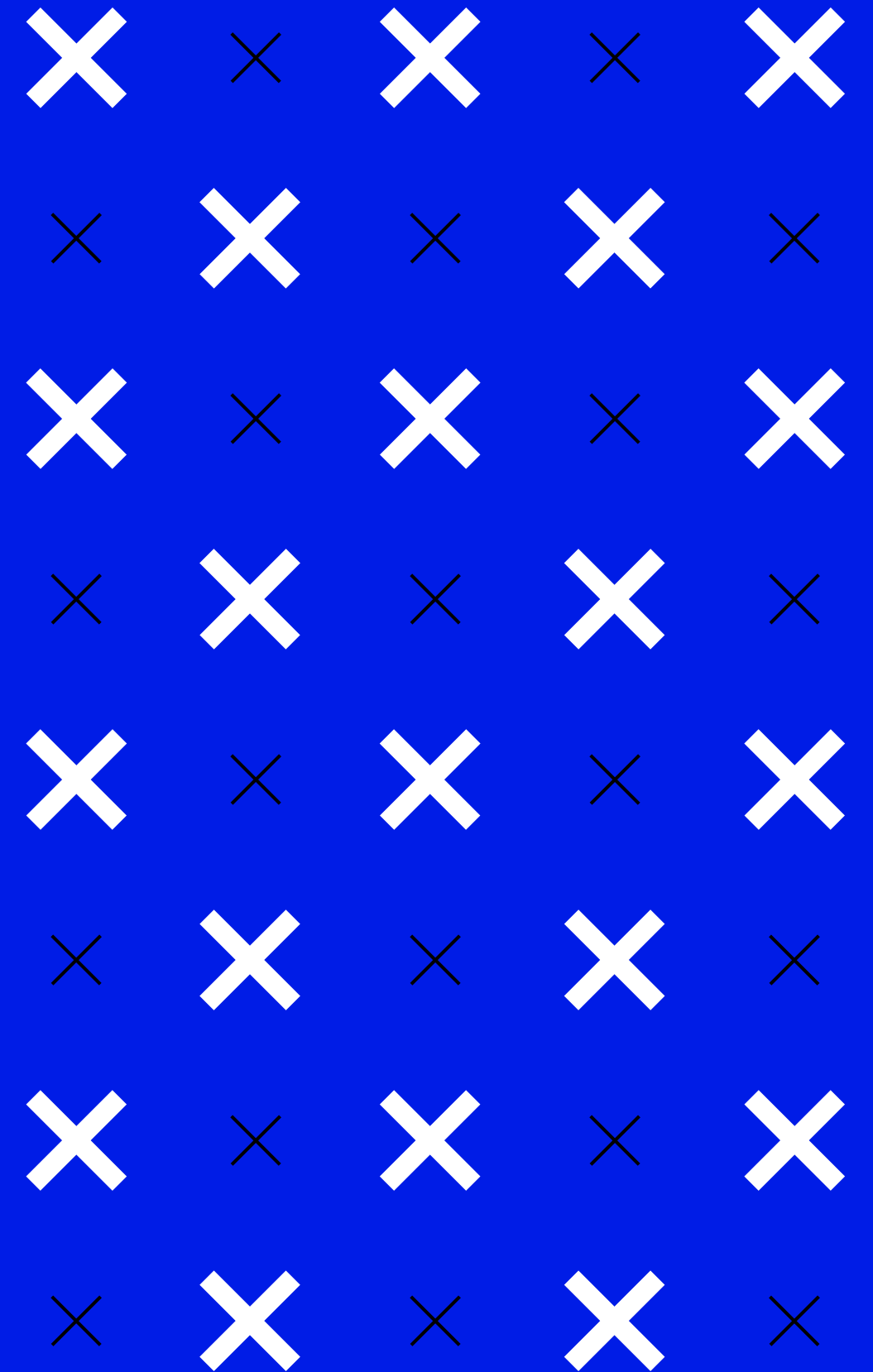


# O que veremos?

- ◆ O que são Markup Languages?
- ◆ Manipulando arquivos JSON
- ◆ Manipulando XML
- ◆ Arquivos YAML
- ◆ Navegando pelo HTML
- ◆ Expressões Regulares
- ◆ O que é um PageRank?

**mentorama.**

**mentorama.**



# O que são Markup Languages?

Python profissional

**mentorama.**

mentorama.

## ◆ Tipos de dados

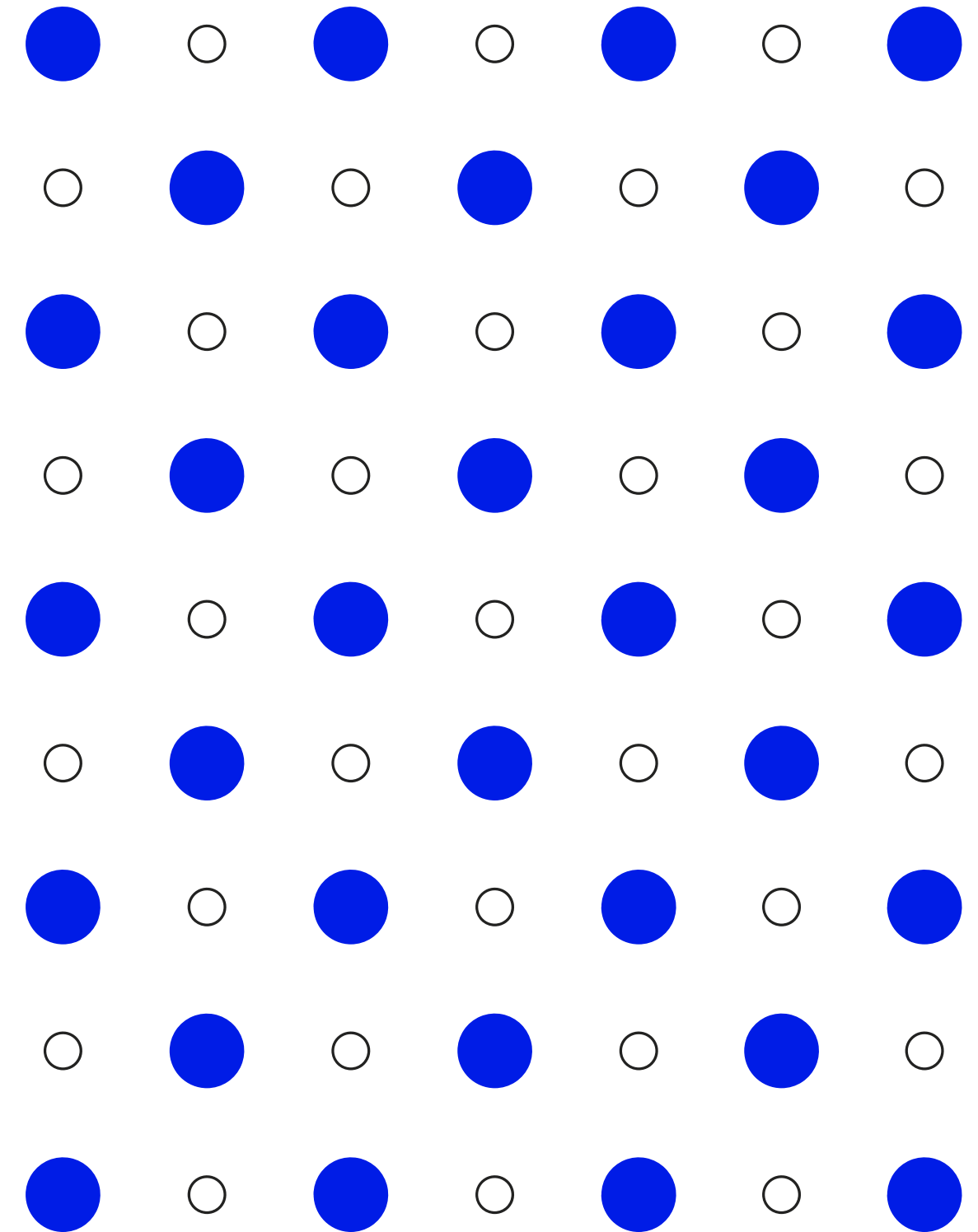
- ◆ Estruturados
- ◆ Semi-estruturados
- ◆ Não estruturados

## ◆ Computadores tem facilidade de trabalhar com dados estruturados

## ◆ Markup Languages não são Linguagens de Programação

**mentorama.**

**mentorama.**



# Manipulando arquivos JSON

Python profissional

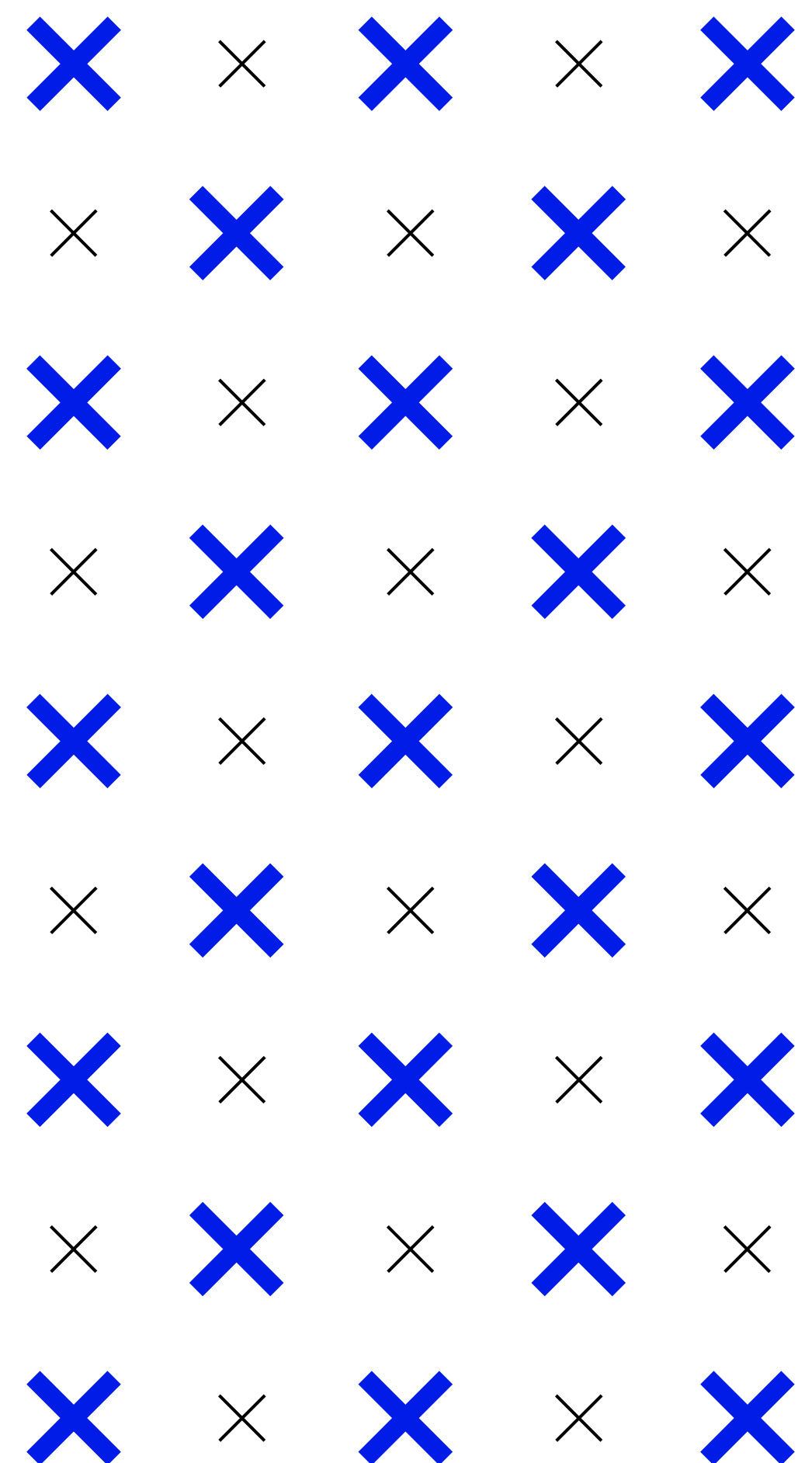
**mentorama.**

mentorama.

- ◆ JavaScript Object Notation (JSON)
- ◆ É um formato de transmissão e armazenamento da informação
- ◆ Onde encontramos o formato JSON?

**mentorama.**

**mentorama.**



Mãos à obra

# Manipulando estruturas JSON com Python

mentorama.

mentorama



# Manipulando XML

Python professional

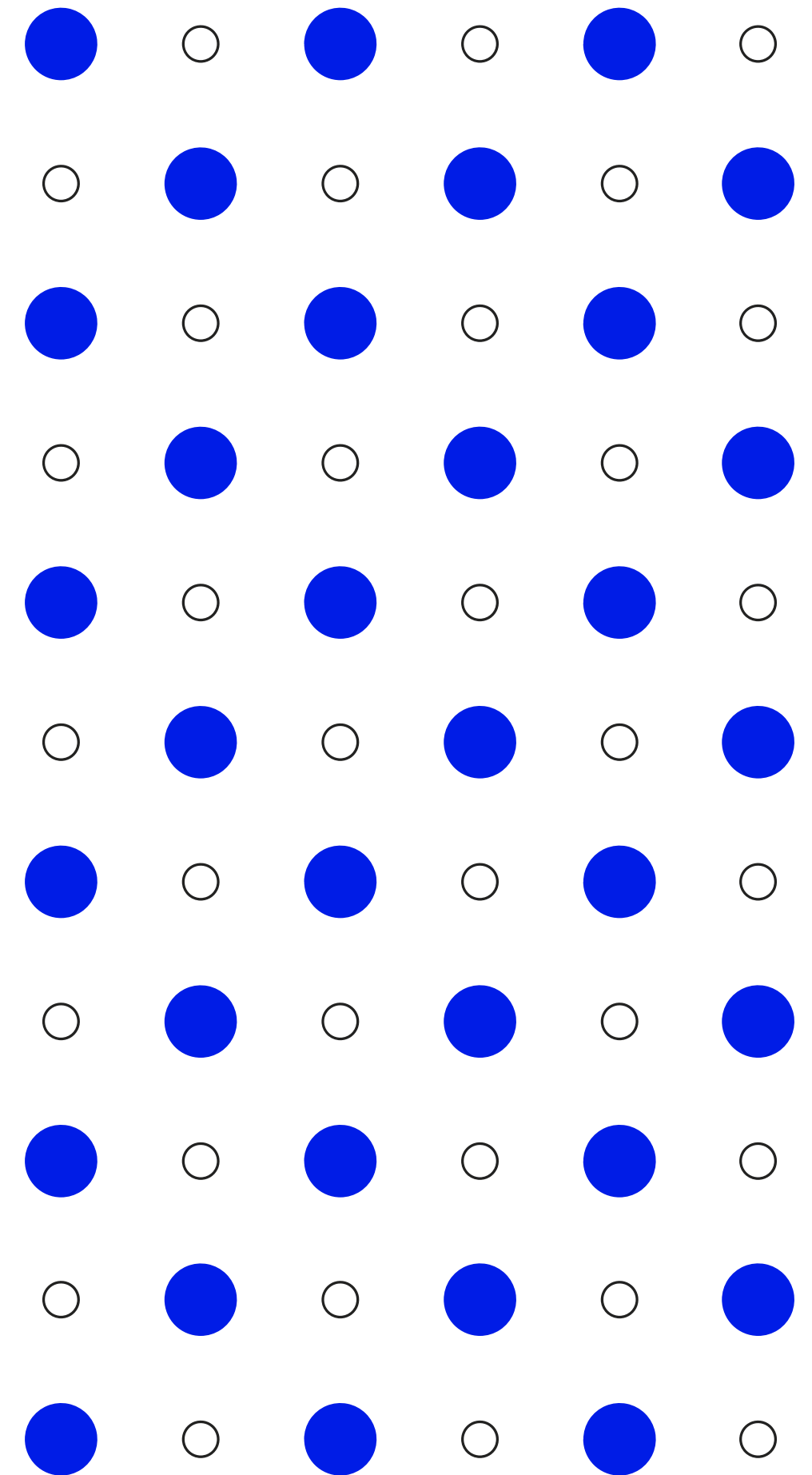
**mentorama.**

mentorama.

- ◆ Extensible Markup Language (XML)
- ◆ Muito utilizado em aplicações financeiras
- ◆ Muito utilizado em arquivos de configuração em frameworks
- ◆ Utiliza <Tags> para delimitar o conteúdo

**mentorama.**

**mentorama.**



Mãos à obra

# Lendo um arquivo XML com Python

mentorama.

mentorama.

# Arquivos YAML

Python professional

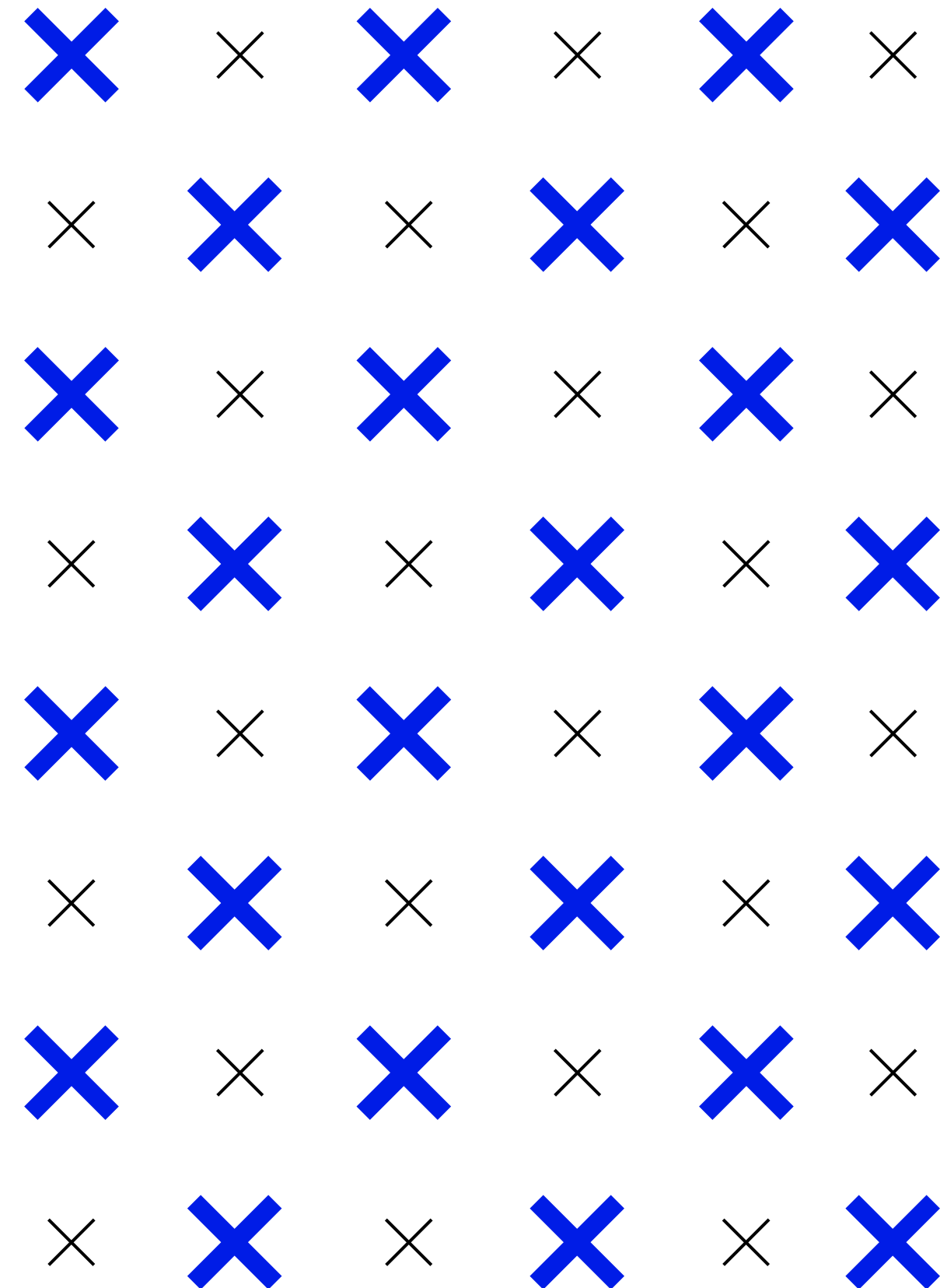
**mentorama.**

mentorama.

- ◆ YAML Ain't Markup Language
- ◆ Considerado uma linguagem de marcação rápida
- ◆ Combina listas, hashes e dados escalares
- ◆ Traz uma formatação mais visível para a leitura humana

**mentorama.**

**mentorama.**



Mãos à obra

# Utilizando Python para leitura e escrita de arquivos YAML

mentorama.

mentorama

# Navegando pelo HTML

Python profissional

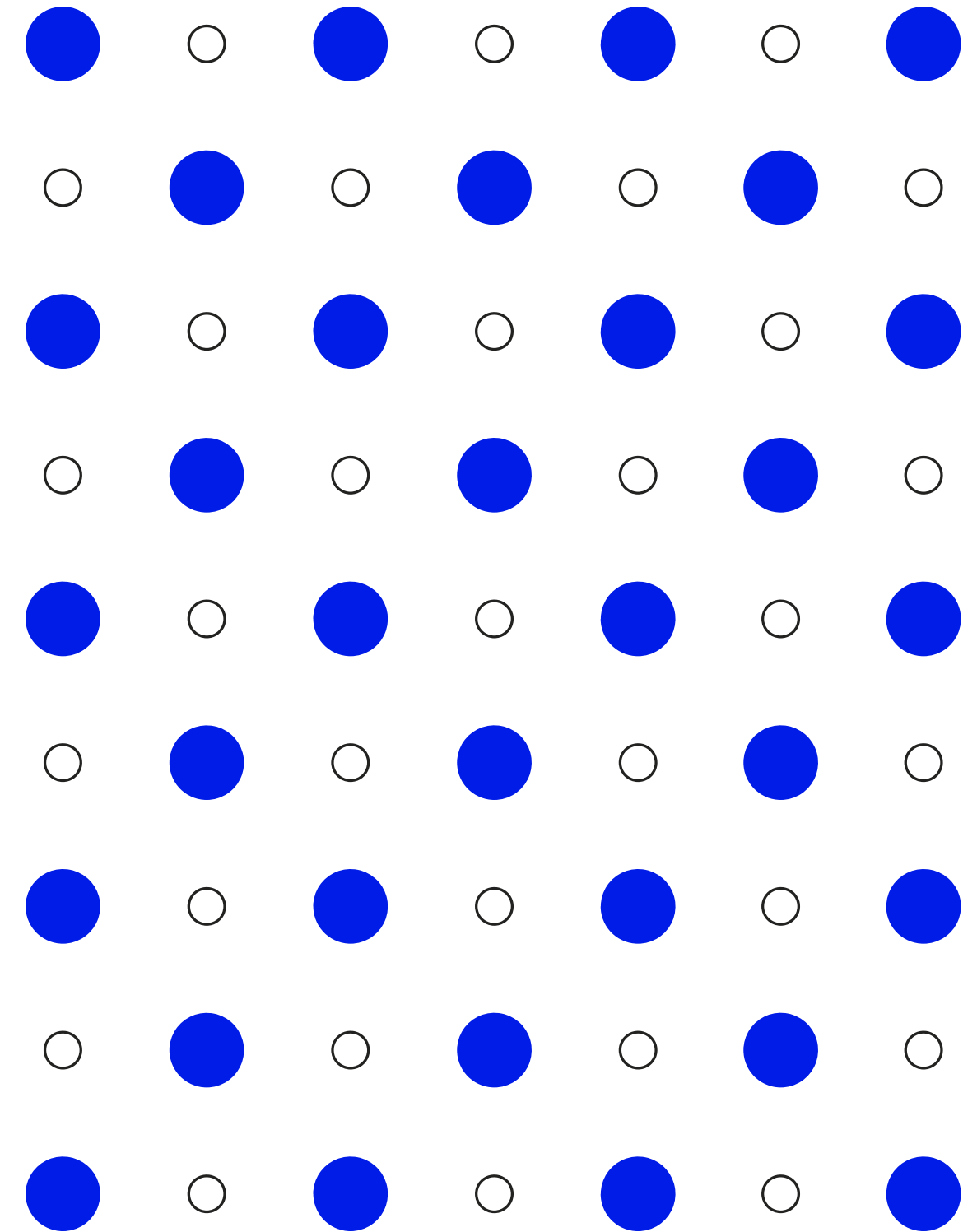
**mentorama.**

mentorama.

- ◆ HyperText Markup Language (HTML)
- ◆ Linguagem de Marcação padrão para criação de conteúdos na Web
- ◆ Versão mais utilizada atualmente é HTML5
- ◆ Funciona com o sistema de tags, assim como o XML

**mentorama.**

**mentorama.**





Mãos à obra

# Utilizando o BeautifulSoup para navegar pelas tags

mentorama.

mentorama.

# Expressões Regulares

Python profissional

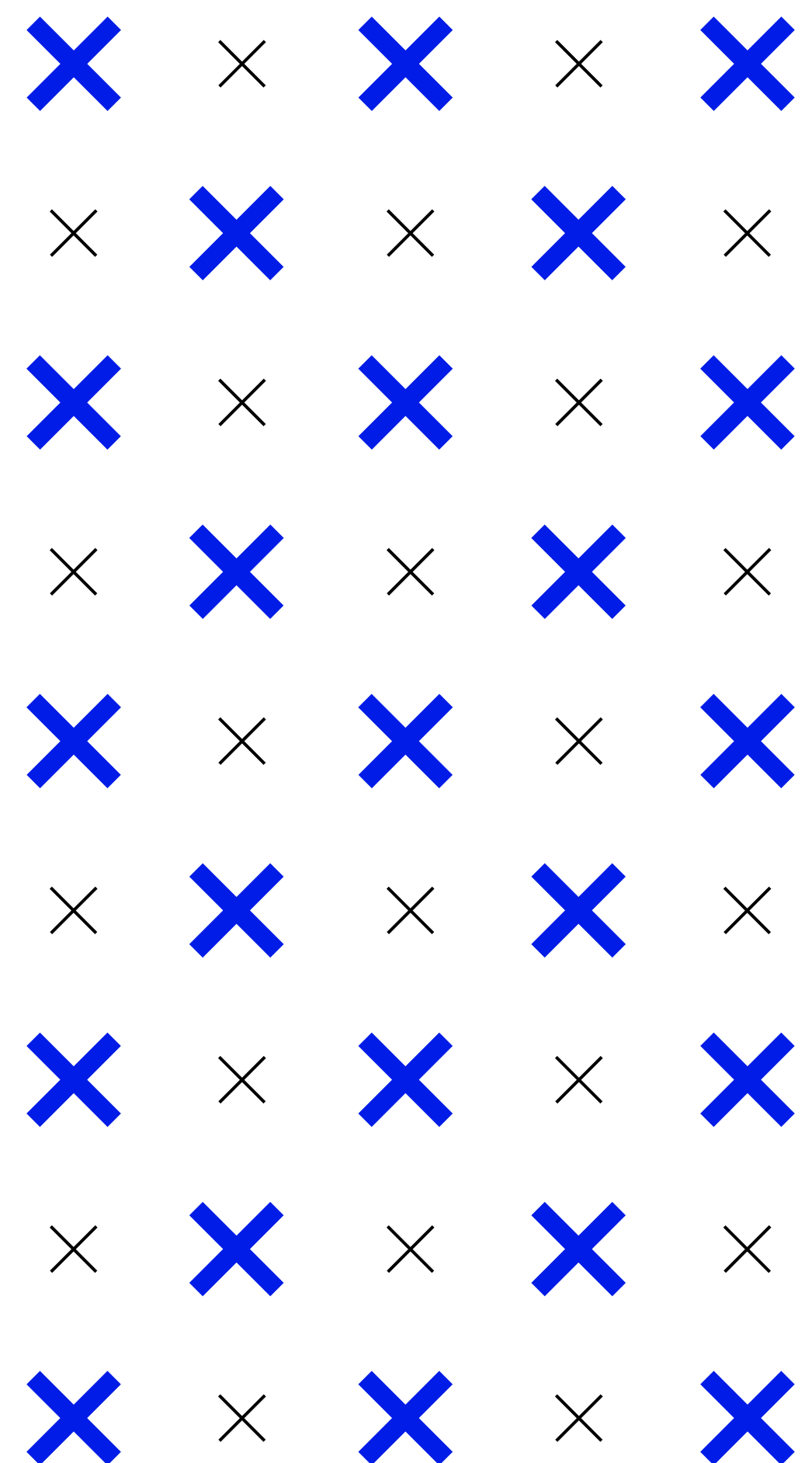
**mentorama.**

mentorama.

- ◆ Também conhecida como Regex, do inglês Regular Expression
- ◆ Estabelece padrões para busca de grupos de caracteres
- ◆ O Regex é amplamente aplicado em:
  - ◇ Sistemas Operacionais
  - ◇ Aplicações Web
  - ◇ Linguagens de Programação
  - ◇ Diversos tipos de sistemas

**mentorama.**

**mentorama.**



Mãos à obra

# Utilizando Expressões Regulares com Python

mentorama.

mentorama.

# O que é um PageRank?

Python profissional

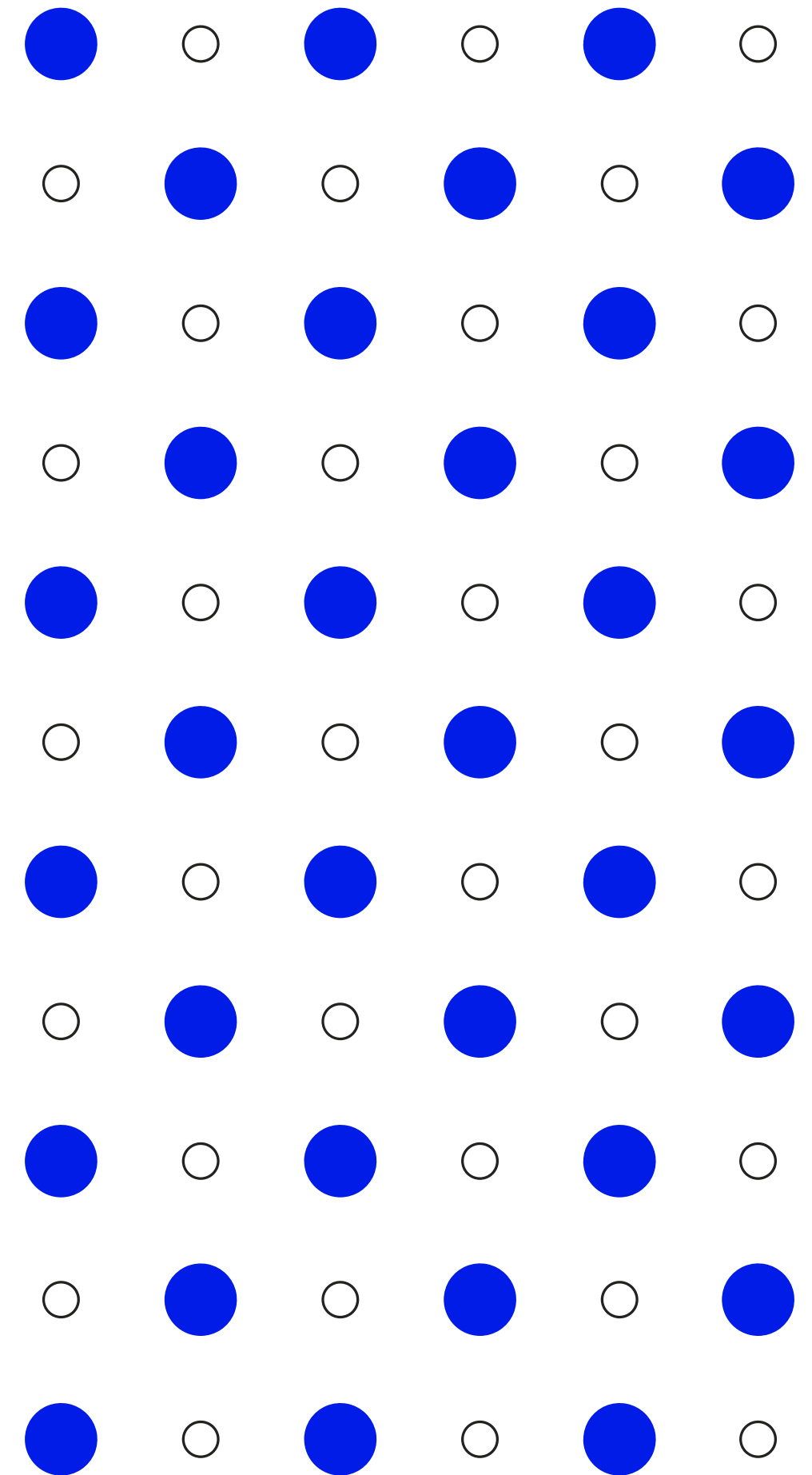
**mentorama.**

mentorama.

- ◆ Artigo científico publicado pelos criadores da Google
- ◆ Utilização das técnicas estudadas neste capítulo e no capítulo anterior
- ◆ Algoritmo de busca por sites na Web e ranqueamento dos sites
- ◆ Por que o Python foi importante para a criação da Google?

**mentorama.**

**mentorama.**



# Tarefa

Desenvolva um Web Scraper para obter um arquivo YAML:

- ◆ Realize a raspagem dos dados da página [www.w3schools.io/file/yaml-sample-example/](http://www.w3schools.io/file/yaml-sample-example/)
- ◆ Extraia o exemplo de arquivo YAML da página
- ◆ Utilize REGEX para extrair os comentários contidos no exemplo e imprima na tela
- ◆ Escreva um arquivo YAML os exemplo extraído do arquivo YAML

**mentorama.**

**mentorama.**

