

# What a Character

Unicode Regular Expressions - Concepts

**Sualeh Fatehi**

# Case Insensitive Patterns

- Use case-insensitive Unicode matching
- Provide case-insensitive flag to regular expression
- Add Unicode flag as well

# Case Insensitive Patterns in Java

```
Pattern pattern = Pattern.compile("σκύλος",  
    Pattern.CASE_INSENSITIVE |  
    Pattern.UNICODE_CASE);  
  
matches = pattern.matcher("ΣΚΥΛΟΣ").matches();
```

# Case Insensitive Patterns in C#

```
Regex patternGreek = new Regex(@"σκύλος",  
    RegexOptions.IgnoreCase |  
    RegexOptions.CultureInvariant);  
  
matches = patternGreek.IsMatch("ΣΚΥΛΟΣ");
```

# Case Insensitive Patterns in JavaScript

```
patternGreek = /σκύλος/iu  
matches = patternGreek.test("ΣΚΥΛΟΣ");
```

# Case Insensitive Patterns in Python

```
pattern_greek = re.compile("σκύλος",  
                             re.IGNORECASE |  
                             re.UNICODE)  
  
matches = bool(pattern_greek.match("ΣΚΥΛΟΣ"))
```

# Simple Pattern for Digits

```
[0-9]
```

- Matches only ASCII digits
- Does not match digits in other languages

# Better Pattern for Digits

```
\d+
```

- Intended to match digits in all languages
- Does not match digits in other languages in Java and JavaScript



# Patterns for Unicode Numbers

```
\p{Nd}
```

- Matches on a Unicode Category
- In this example, matches the Decimal Number Category (Nd)
- Supported by most programming languages

# Code Examples

Slides and all code examples are on GitHub

<https://github.com/sualeh/What-a-Character>

