

# HTML

## 5. HTML AND UNICODE

รวบรวมโดย อ. กิตติพิชญ์ คุปตะวาณิช

### HTML Entities (Revisited)

- การแสดงอักขระพิเศษในเนื้อหาของเอกสาร สามารถทำได้ โดยการใส่รหัสพิเศษ (HTML Entities) ซึ่งมี 2 รูปแบบคือ

- `&entity_name;`
- `&#entity_number;`



Character Escape

Result	Entity Name	Entity Number
	&nbsp;	&#160;
<	&lt;	&#60;
>	&gt;	&#62;
&	&amp;	&#38;
§	&sect;	&#167;
©	&copy;	&#169;
®	&reg;	&#174;
™	&trade;	&#8482;

### What is Unicode?

- คอมพิวเตอร์จัดเก็บตัวอักษรและอักขระอื่น ๆ โดยการกำหนดหมายเลขรหัส (Encoding) ให้สำหรับอักขระแต่ละตัว
- ก่อนหน้าที่ Unicode จะถูกสร้างขึ้น ได้มีระบบ Encoding อยู่หลายร้อยระบบสำหรับการกำหนดหมายเลขเหล่านี้
- ไม่มี Encoding ใดที่มีจำนวนตัวอักขระมากเพียงพอ
- Unicode เป็นมาตรฐานในการกำหนดรหัสสำหรับอักขระในทุก Platform และ ภาษา
- องค์กรที่กำหนดมาตรฐาน Unicode คือ Unicode Consortium

### HTML and Unicode Characters

- การแสดงตัวอักษรในภาษาอื่น ๆ เช่น 你好 (ni hao) รวมถึงอักขระพิเศษเช่นสัญลักษณ์ 卐 หรือ ☯ ในเอกสาร HTML สามารถทำได้โดยใช้วิธี Character Escape (&#\_\_\_;) เช่นกัน
- โดยรหัส Unicode จะเป็นเลขฐาน 16 ในขณะที่ รหัส HTML Entity จะเป็นเลขฐาน 10
- สัญลักษณ์ ☂ สามารถแทนได้ด้วย &#9730; หรือ &#x2602; (เครื่องหมาย x ใช้หน้ารหัสเลขฐาน 16)

Official Name	Glyph	Unicode #	HTML
Black sun with rays	☀	U+2600	&#9728;
Cloud	☁	U+2601	&#9729;
Umbrella	☂	U+2602	&#9730;
Snowman	☃	U+2603	&#9731;
Comet	☄	U+2604	&#9732;

# HTML and Unicode Characters [2]

- นอกจากนี้ยังสามารถทำได้โดยการกำหนดการเข้ารหัส (Encoding) ของไฟล์เอกสารเป็น UTF-8 แล้วพิมพ์เอกสารในภาษานั้น ๆ
- การกำหนดการเข้ารหัสไฟล์เอกสารไม่ใช้การกำหนด Tag `<meta charset="utf-8"/>` ใน `<head>` element
- จำเป็นต้องกำหนดค่า Encoding ของไฟล์เอกสารผ่าน Editor ที่ใช้สร้างเอกสาร

204202: Information Technology II

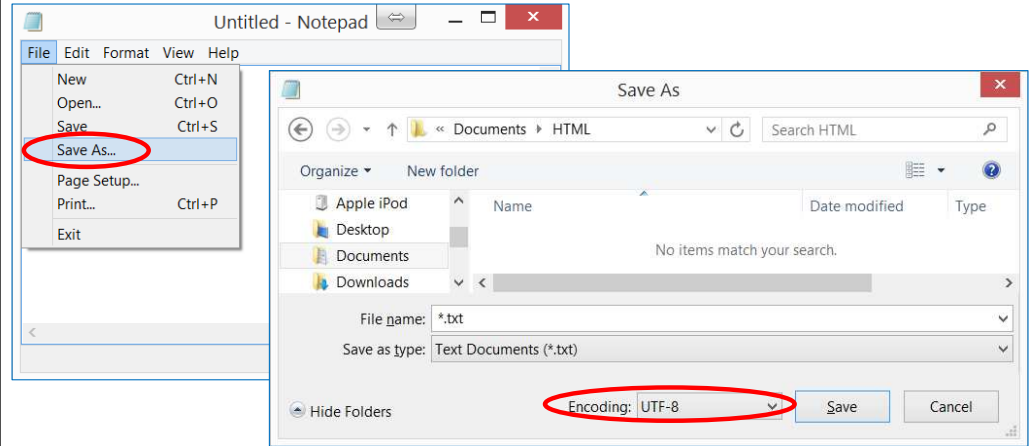
5

Ref: <http://www.w3.org/International/questions/qa-escapes>

# HTML and Unicode Characters [2]

## Notepad Application

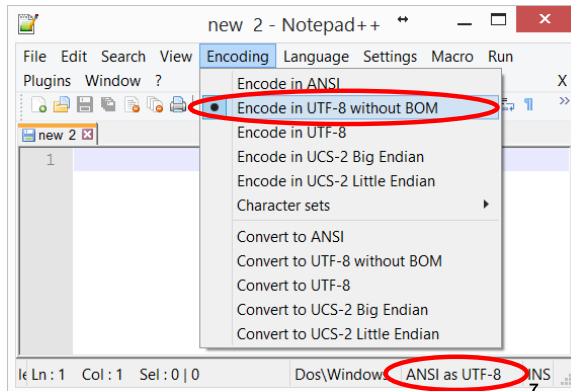
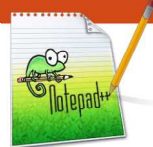
- ทำได้โดยการเลือก Encoding เป็น UTF-8 ขณะ save เอกสาร



# HTML and Unicode Characters [3]

## Notepad++ Application

- ทำได้โดยการเลือก "Encode in UTF-8 without BOM" ผ่าน Menu Encoding
- ใน status bar จะแสดงผลเป็น ANSI as UTF-8
- ตามมาตรฐาน Unicode ไม่แนะนำให้ใช้ BOM (Byte Order Mark) กับ UTF-8



Ref: <http://www.unicode.org/versions/Unicode5.0.0/ch02.pdf>

# Typing Unicode Characters

- ใน Application บน MS Windows การพิมพ์ อักขระจากรหัส Unicode ทำได้โดยการ กดปุ่ม Alt ค้างไว้ ตามด้วยรหัสเลขฐาน 10 แล้วปล่อย เช่น Alt 9730 จะได้ ☂
- Application นั้น ๆ ต้องสามารถรับ input แบบ Unicode ได้เช่น MS Office หรือ WordPad

Official Name	Glyph	Unicode #	HTML
Black sun with rays	☀	U+2600	&#9728;
Cloud	☁	U+2601	&#9729;
Umbrella	☂	U+2602	&#9730;
Snowman	☃	U+2603	&#9731;
Comet	☄	U+2604	&#9732;



204202: Information Technology II

8

# References

- **Unicode คืออะไร**

<http://www.unicode.org/standard/translations/thai.html>

- **W3C: Using character escapes in markup and CSS**

<http://www.w3.org/International/questions/qa-escapes>

- **Wikipedia: List of Unicode characters**

[http://en.wikipedia.org/wiki/List\\_of\\_Unicode\\_characters](http://en.wikipedia.org/wiki/List_of_Unicode_characters)

- **Wikipedia: Miscellaneous Symbols**

[http://en.wikipedia.org/wiki/Miscellaneous\\_Symbols](http://en.wikipedia.org/wiki/Miscellaneous_Symbols)