

占쏙올 占쏙올

최원우

UPnL 워크샵, 2018-01-21

占속엍 占속엍



占속엍占속엍 占속엍占속엍 占속엍占속엍 占속엍=

Sun Java Enterprise System 2003Q4 占속엍치 占속엍占속엍?

1 占속엍

占속엍占속엍

占속엍 占속엍占속엍占속엍 Sun Java占속엍 Enterprise System 占속엍 Java Enterprise System 占속엍치 占속엍占속엍占속엍占속엍占속엍 占속엍占속엍 占속엍占속엍를 f占속엍占속엍
니담엍.

占속엍 占속엍占속엍 占속엍占속엍: 占속엍=占속엍 占속엍占속엍 4求占속엍.

- [Java Enterprise System](#)
- [Java Enterprise System 占속엍치 占속엍占속엍占속엍占속엍 占속엍디엍 占속엍占속엍](#)
- [Java Enterprise System 占속엍占속엍트 占속엍占속엍를 占속엍占속엍엍 占속엍占속엍](#)

Java Enterprise System


Java Enterprise System: Sun 占속엍 占속엍占속엍占속엍 f동; 占속엍占속엍 占속엍占속엍트 占속엍占속엍 占속엍占속엍8占속엍 占속엍占속엍占속엍 占속엍삼엍占속엍 占속엍占속엍占속엍占속엍占속엍占속엍占속엍? 占속엍 占속엍占속엍占속엍; 占속엍占속엍求占속엍 占속엍 占속엍엍占속엍 占속엍占속엍 占속엍占속엍트 占속엍占속엍를 f占속엍
占속엍니담엍. 占속엍 占속엍占속엍 占속엍占속엍占속엍占속엍 SPARC 占속엍엍占속엍占속엍 Sun Solaris占속엍 8 占속엍 9 占속엍 占속엍 占속엍 X86 占속엍엍占속엍占속엍
Solaris 9 占속엍 占속엍占속엍占속엍 占속엍占속엍占속엍 占속엍 占속엍 4求占속엍.

Java Enterprise System 占속엍 占속엍 占속엍占속엍; 占속엍占속엍占속엍占속엍 占속엍= 占속엍; 占속엍보占속엍시웍엍.

- [Java Enterprise System 占속엍 占속엍a](#)
- [占속엍占속엍占속엍占속엍占속엍占속엍 占속엍트 占속엍크 占속엍占속엍](#)

<https://docs.oracle.com/cd/E19199-01/817-4243-10/overview.html>

왜 생기는가

- 어떤 이유로 인코딩이 잘못되어 로 대체됨
(U+FFFD REPLACEMENT CHARACTER)
- 이걸 UTF-8로 인코딩하면 EF BF BD
- 2개 연속되면 EFBF BDEF BFBD
- 이걸 Windows-949(EUC-KR)로 해석하면 占쏙옷
- → 잘못돼도 한참 잘못됐을 때 생김

발생 가능한 시나리오

- UTF-8이 아닌 웹 문서를 억지로 UTF-8로 해석
 - 해석 불가능하므로 대체 문자가 대신 들어감
- 서버가 따로 인코딩 정보를 전달하지 않음
- 브라우저는 일단 Windows-949로 해석함
- → **占쏙옷**이 잔뜩 찍혀나옴

직접 만들어 보기 – Python 3

- 방금 전 시나리오를 그대로 코드로 옮기면 됨
- decode 메서드의 'replace' 옵션 활용
 - 해석할 수 없는 경우 에러 대신 대체 문자 삽입

```
def break_949(s):  
    return s \  
        .encode('cp949') \  
        .decode('utf-8', 'replace') \  
        .encode('utf-8') \  
        .decode('cp949', 'replace')
```


실행 결과 - Python 3

```
>>> from broken import break_949  
>>> break_949( '가나다라' )  
' 占쑈엵 占쑈엵 占쌔띤엵 '
```


占색뎡엣

```
>>> '다라'.encode('cp949')  
b'\xb4\xd9\xb6\xf3'
```

?	2 bytes	cont.	4 bytes	?
10110100	11011001	10110110	11110011	[EOF]
Error!	11001	110110		Error!
→ U+FFFD	→ U+0676		→ U+FFFD	
	ARABIC LETTER			
	HIGH HAMZA WAW			

占쌔띤엿

U+FFFD

U+0676

REPLACEMENT CHARACTER

ARABIC LETTER HIGH HAMZA WAW

11101111 10111111 10111101 11011001 10110110

EFBF

BDD9

B6

U+FFFD

REPLACEMENT CHARACTER

11101111 10111111 10111101

EF

BFBD

```
>>> b'\xef\xbf\xbd\xd9\xb6\xef\xbf\xbd' \
...     .decode('cp949', 'replace')
'占쌔띤엿'
```


직접 만들어 보기 – Rust

- rust-encoding

```
extern crate encoding;
```

```
use encoding::all::WINDOWS_949;
```

```
use encoding::{Encoding, EncoderTrap, DecoderTrap};
```

```
fn break_encoding(s: &str) → String {  
    let first_949 = WINDOWS_949.encode(s, EncoderTrap::Replace).unwrap();  
    let second_utf8 = String::from_utf8_lossy(&first_euc_kr);  
    WINDOWS_949.decode(second_utf8.as_bytes(), DecoderTrap::Replace).unwrap()  
}
```


실행 결과 – Rust

```
fn main() {  
    println!("{}", break_encoding("가나다라"));  
}
```

```
$ cargo run  
    Finished dev [unoptimized + debuginfo] target(s) in 0.0 secs  
    Running `target/debug/broken-949`  
占쑝엎占쑝엎占쌔띐엎
```


완전 무작위 생성 - Python

- ‘가’부터 ‘힉’까지를 여러 번 골라서 문자열 생성
- 그대로 깨뜨림

```
def break_random(n):  
    randstr = ''.join([  
        chr(random.randrange(0xac00, 0xd7a4)) for _ in range(0, n)  
    ])  
    return break_949(randstr)
```

```
>>> from broken import break_random  
>>> break_random(10)  
' 𐄀쑈엎 𐄀씸쟈엎 𐄀쑈엎 𐄀뽏 𐄀뽏 𐄀쑈엎 𐄀뽏 𐄀쑈엎 '
```


완전 무작위 생성 – Rust

- rand

```
fn random_break(n: usize) → String {  
    use rand::distributions::IndependentSample;  
    let mut rand = rand::thread_rng();  
    let hangul_range = rand::distributions::Range::new(0xac00, 0xd7a4);  
    let random_str = (0..n)  
        .into_iter()  
        .map(|_| unsafe {  
            std::char::from_u32_unchecked(hangul_range.ind_sample(&mut rand))  
        })  
        .collect::<String>();  
    break_encoding(&random_str)  
}
```


완전 무작위 생성 – Rust

```
fn main() {  
    println!("{}", random_break(10));  
}
```

```
$ cargo run  
    Finished dev [unoptimized + debuginfo] target(s) in 0.0 secs  
    Running `target/debug/broken-949`  
占뽕占뽕占신틈占쓱옎占뽕占뽕占뽕占쓱옎占뽕
```


이걸 어디에 써요?

- 글자를 깨뜨릴 필요가 있을 때

```
$ gtext = glitchtext(20)
y 1s "그래도 이해해주신다니 정말 감{nw}"
# 생략
y glitch "그래도 이해해주신다니 정말 감{fast}[gtext] [gtext][gtext]{nw}"
```

그래도 이해해주신다니 정말 감¢ ¨řÛŕ;ŕ®ŒèÉěćŁÀùĴŎŴg¢ ¨řÛŕ;ŕ®ŒèÉěćŁÀùĴŎŴg¢ ¨řÛŕ;ŕ®ŒèÉěćŁÀùĴŎŴg

그래도 이해해주신다니 정말 감占쑈엎冤눅엎책t占쑈엎 占쑈엎冤눅엎책t占쑈엎 占쑈엎冤눅엎책t占쑈엎

- 이 발표를 하게 된 이유이다