

Convert to ASCII  
ŠCEŽšœžŸŸμÀáâãäåæçè  
↓  
SOZsozYYuAAAAAAACE

Upper case  
abCDeFg → ABCDEFG

Decode HTML  
&OACUTE; → O  
<BR /> → \_

Remove non-alphanumeric  
,,.?/\*%^#@ → \_

Umlaut harmonization  
ue, ae, oe → u, a, o

Add spaces  
ABCDEFG → \_ABCDEFG\_

Harmonize conjunctions  
+ → &  
AND → &  
ET → &  
UND → &

Standardize  
ACTIEN GESELLSCHAFT → AG  
ACTIENGESELLSCHAFT → AG  
AKTIEN GESELLSCHAFT → AG  
AKTIENGESELLSCHAFT → AG

Remove common words  
THE → \_  
GMBH → \_  
INC → \_

Trim spaces  
\_ABC\_ \_DEF\_ \_G\_ → ABC\_DEF\_G