



# pears.GetTextFromHTML



Generated schema reference. Regenerate this page from the SQL unload; keep hand-maintained business notes in the narrative namespace.

## Original SQL

```
CREATE FUNCTION "pears"."GetTextFromHTML" (  
  /* Application Maintained Function / Procedure - DO NOT EDIT*/  
  IN "htmlText" long VARCHAR )  
RETURNS long VARCHAR  
BEGIN  
  DECLARE "plainText" long VARCHAR;  
  DECLARE "tagStart" INTEGER;  
  DECLARE "tagEnd" INTEGER;  
  DECLARE "tagLength" INTEGER;  
  DECLARE "tagType" CHAR(20);  
  DECLARE "endOfTagName" INTEGER;  
  SET "plainText" = "htmlText";  
  IF "locate"("plainText", '<HEAD>', 1) = 0 THEN  
    RETURN "plainText"  
  END IF;  
  SET "plainText" = "right"("plainText", "length"("plainText") -  
"charindex"('</HEAD>', "upper"("plainText")) - 6);  
  SET "plainText" = "replace"("plainText", "char"(13), '');  
  SET "plainText" = "replace"("plainText", "char"(10), '');  
  SET "plainText" = "replace"("plainText", ',', "char"(96));  
  SET "plainText" = "replace"("plainText", '@', "char"(96));  
  while "locate"("plainText", '<', 1) > 0 loop  
    SET "tagStart" = "charindex"('<', "plainText");  
    SET "tagEnd" = "charindex"('>', "plainText");  
    // make tag TYPE always be IN UPPER CASE, AND compensate FOR the <  
things  
    SET "tagType" = "substring"("upper"("plainText"), "tagStart"+1, "tagEnd"-  
"tagStart"-1);  
    SET "tagLength" = "tagEnd" - "tagStart";  
    // CHECK TO see IF there are spaces IN the name (there are attributes).  
    // IF there are GET rid OF them  
    SET "endOfTagName" = "charindex"(' ', "tagType");  
    IF "endOfTagName" > 0 THEN  
      SET "tagType" = "substring"("tagType", 1, "endOfTagName")  
    END IF;
```



```
CASE "tagType"
WHEN '/DIV' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/P' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H1' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H2' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H3' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H4' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H5' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/H6' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/UL' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/OL' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN 'BR' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/TABLE' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN 'HR' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/ARTICLE' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/ASIDE' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/BLOCKQUOTE' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/FIGCAPTION' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/FIGURE' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/FOOTER' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/FORM' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/HEADER' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/MENU' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/NAV' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
WHEN '/PRE' THEN SET "plainText" =
```



```
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
  WHEN '/SECTION' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10))
  WHEN '/SUMMARY' THEN SET "plainText" =
"stuff"("plainText","tagStart","tagLength"+1,"char"(13)+"char"(10)) ELSE
  SET "plainText" = "stuff"("plainText","tagStart","tagLength"+1,NULL)
END CASE
END loop;
SET "plainText" = "replace"("plainText",'&nbsp;',' ');
SET "plainText" = "replace"("plainText",'&','&');
SET "plainText" = "replace"("plainText",'&gt;','>');
SET "plainText" = "replace"("plainText",'&lt;','<');
SET "plainText" = "replace"("plainText",'&quot;','char"(34));
SET "plainText" = "replace"("plainText",'&apos;','char"(39));
SET "plainText" = "replace"("plainText",'&cent;','char"(163));
SET "plainText" = "replace"("plainText",'&pound;','char"(163));
SET "plainText" = "replace"("plainText",'&euro;','char"(128));
SET "plainText" = "replace"("plainText",'&yen;','char"(165));
SET "plainText" = "replace"("plainText",'&copy;','char"(169));
SET "plainText" = "replace"("plainText",'&reg;','char"(174));
IF("isnull"("plainText",'') = '') AND("isnull"("htmltext",'') <> '') THEN
  SET "plainText" = 'Could not parse following
HTML: '+'char"(13)+"char"(10)+"htmlText"
END IF;
RETURN "plainText"
END
```

From:

<https://iqxusers.co.uk/iqxhelp/> - iqx

Permanent link:

[https://iqxusers.co.uk/iqxhelp/doku.php?id=database:functions:pears\\_gettextfromhtml](https://iqxusers.co.uk/iqxhelp/doku.php?id=database:functions:pears_gettextfromhtml)

Last update: 2026/08/07 19:24

