u2d_msa_sdk Module

`.utils.htmlutils`

Attributes

__version__ module-attribute

```
__version__ = '0.0.8'
```

Functions

sanitize async

```
sanitize(dirty_html: Any) -> Optional[str]
```

> Clean/Sanitize HTML using lxml.html.clean.Cleaner
>
> Cleans the following:
>
> - Removes any `<meta>` tags
> - Removes any embedded objects (flash, iframes)
> - Removes any `<link>` tags
> - Removes any style tags.
> - Removes any processing instructions.
> - Removes any style attributes. Defaults to the value of the `style` option.
> - Removes any `<script>` tags.
> - Removes any Javascript, like an `onclick` attribute. Also removes stylesheets as they could contain Javascript.
> - Removes any comments.
> - Removes any frame-related tags
> - Removes any form tags
> - Removes Tags that aren't *wrong*, but are annoying. `<blink>` and `<marquee>`

- Remove any tags that aren't standard parts of HTML.
- Remove any attributes which are not frozenset(['src', 'color', 'href', 'title', 'class', 'name', 'id']),
- Remove Tags ('span', 'font', 'div'), their content will get pulled up into the parent tag.

PARAMETER	DESCRIPTION
`dirty_html`	Any, usually a html str **TYPE:** `Any`

RETURNS	DESCRIPTION
`clean_html`	Optional[str] cleaned html **TYPE:** `Optional[str]`

Last update: September 13, 2022

Created: September 13, 2022