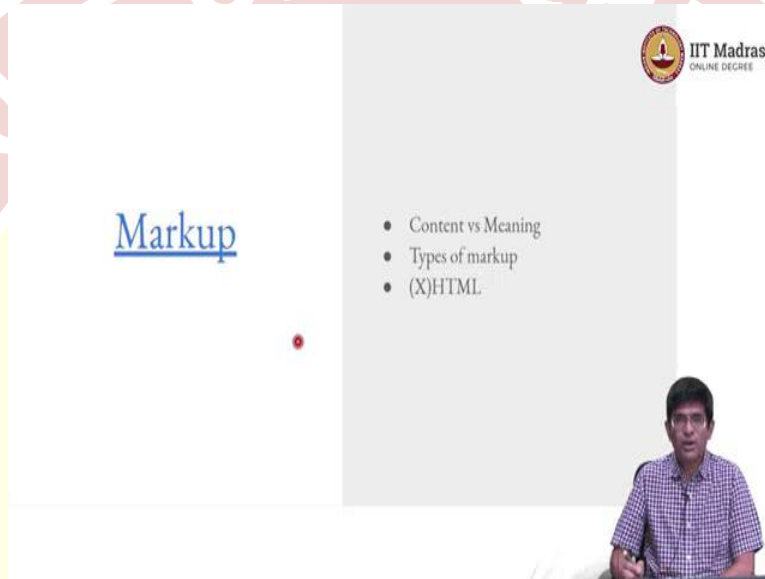


**IIT Madras**  
ONLINE DEGREE

**Modern Application Development - I**  
**Professor. Nitin Chandrachoodan**  
**Department of Electric Engineering**  
**Indian Institute of Technology, Madras**  
**What is Markup?**

Hello, everyone, and welcome to this course on Modern Application Development.

(Refer Slide Time: 0:16)



So, now let us get to the question of what do we mean by markup? And what we need to understand over here is the difference between raw content versus what that content means. We will also look at some different types of markup. And at variants, I am going to call it X HTML. But that is basically because ultimately we are looking at the Hypertext Markup Language, HTML. There are many variants around it that we need to understand here.

(Refer Slide Time: 0:44)


Content

 IIT Madras  
ONLINE DEGREE

Markup What is markup? Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand.




Markup

 IIT Madras  
ONLINE DEGREE

Title      Heading level 1

Markup What is markup? Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand.

Insert para break



सिद्धिर्भवति कर्मजा

# Markup

## What is markup?

Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed.

Markup is very useful to make the display of text clear and easy to understand.



So, let us look at this document that I have up on the screen here. This is what I mean by a raw content. Essentially, what you can see is markup. What is markup? Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand. I can read it out. But it is not very clear what is important. What was the beginning? What was the ending? What was somewhere else?

Already over here, you can see that there is some kind of markup. For example, I used a capital letter at the beginning of the sentence. I used a punctuation mark, a question mark. I use a full stop over here, all of these things. And after the full stop, I have another capital letter. Now, these things themselves are a form of markup. Why are they there? Because it makes it easier for humans to understand what is done with, how to interpret the sentences.

But let us say I see a document like this, and I say, look, this is not very clear, can you rearrange this a little bit. And I start making some drawings on it and send it back to the person who wrote it, saying, look, this part over here, this first word, this markup, that should be the title. So, please set it so that it becomes clear that this is a title.

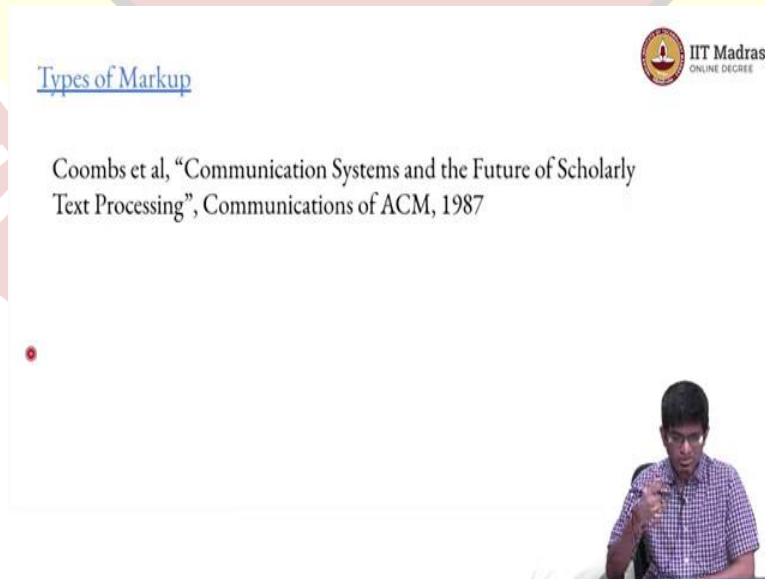
Then you have this question over here, what is markup? Rather than just having it as a question the flow of the text, why do not you make it a heading, so that it becomes a bit more easy to understand what this part of the writing is about. Not just that, insert a paragraph break over here, because it helps to sort of make this a bit more clear and easy to read.

What I have done over here, all these things that are shown is essentially what markup was. It literally meant that I would take a document printed on a piece of paper or written on a piece of paper and scribble on it. And say, this part is not clear, this should be rewritten, this should be done something else. Those are, in some sense, the editorial comments. Some of them could just be like mistakes that are being pointed out or suggestions for rewriting that is different. So, that should actually be taken as how do I actually modify the content itself.

But what I have done over here, I am not asking you to modify the content in any way, I am only asking you to modify presentation. And if all of these changes that I have marked up over here were implemented, the document would look something like this. This now becomes the title. The part that I said should be a heading is now clear. It stands out from the rest of the text. It is in a different font. It is slightly bolder. It is larger font size. Not as big as the title, but still big enough that it clearly shows that it is a different sort of meaningful part of the text.

And the paragraph break over here makes it very clear that something has now shifted down, that there is a difference. At this point is in one paragraph and start another paragraph over here. In terms of reading and how the ideas are interpreted, it becomes a little bit easier to understand what is happening. So, this is the idea behind markup. You want to be able to embed some kind of information into the content of the text such that it gets displayed better.

(Refer Slide Time: 4:06)



The slide features a large, faint watermark of the IIT Madras logo in the background. The slide content includes the title 'Types of Markup' in blue text, the IIT Madras logo in the top right corner, and a citation: 'Coombs et al, "Communication Systems and the Future of Scholarly Text Processing", Communications of ACM, 1987'. A small red dot is visible on the left side of the slide. In the bottom right corner, there is a small video inset showing a man in a checkered shirt speaking.

Types of Markup

IIT Madras  
ONLINE DEGREE

Coombs et al, "Communication Systems and the Future of Scholarly Text Processing", Communications of ACM, 1987

Now, markup by itself has been around for a very long time. Those of you who have used the LaTeX text processing system know that it has its origins in the 70s. LaTeX is a form of markup. But the idea of markup itself is fairly old, at least from the 70s, if not earlier. There was a paper published in the communications of the ACM, the Association for Computing Machinery in 1987, which interestingly talks about communication systems and the future of scholarly text processing, essentially talking about what kind of different, how should scholarly documents be written.

And the reason why they are sort of focusing on scholarly documents is because this actually makes a difference. I mean, if we want a machine or some kind of automated process to be able to go through a document and find out which are the important parts, there has to be something more than just the text of the document which allows the machine to understand where to focus.

For those who have read scientific papers, you would know that there is an introduction, there is an abstract, there is a result section, conclusions, each of those are sort of logical sections of your document, which you expect to see, which makes it easy to understand what the whole document is about.

(Refer Slide Time: 5:28)

**Types of Markup**

- Presentational
  - WYSIWYG: directly format output and display
  - Embed codes not part of regular text, specific to the editor
- Procedural
  - Details on how to display:
    - change font to large, bold
    - skip 2 lines, indent 4 columns
- Descriptive
  - This is a <title>, this is a <heading>, this is a <paragraph>

Coombs et al, "Communication Systems and the Future of Scholarly Text Processing", Communications of ACM, 1987

IIT Madras  
ONLINE DEGREE

So, in this paper, they essentially discussed a few different types of markup. One of them is what is called presentational markup. And presentational markup usually goes along with this term WYSIWYG, which you may or may not have heard of. It was a big deal along a few years ago,



well, probably 30 or 40 years ago. It is likely that you still come across this term once in a while, although it is not so common anymore. What it stands for is, what you see is what you get.

And why is that a big deal, because text processors in the old days did not have a, what you see is what you get kind of behavior. You would do your markup in a certain way using even an editor like WordStar or WordPerfect. And you could then get translated into something else when you actually tried to print it out. Microsoft Word was a big jump compared to that. What it meant was that, as and when you typed it, it would automatically show you the thing in a larger font, centered, bold, all of those things will be there for you to see. So, you would see exactly what you will finally get when you print it out on paper.

In this thing, what they do is very often they embed codes which are not really part of regular text. So, if I try reading through a Word document, for example, the raw bytes corresponding to word document, the bytes would not make any sense. They make sense to the particular editor. So, they are specific to the particular editor or word processor that you are using, but are difficult for a human to understand, but are good enough. In terms of presentation and markup, it probably tells you that, look, this part should be bold, this part should be italics, this part should be larger font size, all of those things can be specified directly as part of the presentation.

Now, the next type of markup that they define in this paper is something called procedural markup, which sort of goes into the details on how to display something. And in that, if we go back to this page, for example, the how to display would probably have some small code in front of or somewhere near this word markup, which says switch over to centered text, large font size, let us say 24 point font size, Arial font, and display the text. Then leave two blank lines, change font size to 16, once again, Arial, print what is markup, leave one line, change font size to 12, change font type to Times New Roman, print the next text, leave one blank line.

So, in other words, every single step is sort of specified as part of the markup. Now, is this a good thing to do? It is a different way of handling. The presentation sort of, presentational markup just embedded it directly in the text. Procedural gives you more instructions so that the interpreter that is going through and displaying it needs to be sort of less intelligent. It just needs to follow instructions. But the most important in some sense form of markup is neither of these. It is something else that is called descriptive markup.

And in descriptive markup, we take a completely, well, not completely different approach, but the focus shifts. It is no longer on how the thing looks, but on what it means. So, for example, I would mark up something by mentioning this is a title. And how should a title be displayed is now determined by the tool. The tool has some idea of how titles are to be displayed. Something else would be marked as a heading. Something else would be marked as a paragraph.

Why is this more useful than the original presentational or procedural approaches, because this is neutral to that. Even 40 years later, when we look at it, if I have something which is clearly marked up in terms of a title, a heading and a paragraph, I can automatically go and search through titles. I do not need to go around looking through the text and finding out which one is in bold font or large font sizes.

I know exactly which is the title. So, it has, in other words, specified the meaning of the content rather than how it should be displayed. In some sense that is more useful. You can separate out and say look, now that I have given you the meaning, you decide how you want to display it.

(Refer Slide Time: 10:09)

### Examples

- MS Word, Google Docs etc:
  - User interface focused on "appearance", not meaning
  - WYSIWYG: direct control over styling
  - Often leads to complex formatting and loss of inherent meaning
- LaTeX, HTML (general "ML")
  - Focus on meaning
  - More complex to write and edit, not WYSIWYG in general



So, things like Microsoft Word, Google Docs, and so on are more focused on the WYSIWYG, what you see is what you get. They have direct control over styling. This very often leads to complex formatting and one of the places where you will find that this is actually problematic is, it is possible to write HTML using a Word document and then export to HTML. In most cases that is really difficult to read after that.



The HTML that is generated by a Word document in a lot of cases turns out to be extremely difficult to read, especially if you have some extra styling, some parts in italics or some other special characters or anything else over there. It adds in a lot of extra tags and things which just make it difficult to understand what are you conceptually trying to convey here.

Now, things like LaTeX and even HTML, Hypertext Markup Language, and in general, many of the other so called SGML, XML, all of those markup languages, they tend to focus on the meaning of the document. That is to say, rather than worrying about whether something is in bold font or in large font size, you just specify whether this is a heading, this is an abstract, this is a section, this is a paragraph and somebody else has to decide on how to write it.

In general, they are not WYSIWYG. They are not, what you see is what you get. And they are also more complex to write and to edit. In some cases, there are sort of halfway editors that try to do a good job of, a reasonable job of showing you what the output would look like. But in most cases, they are generally harder to work with than MS Word, which is why people tend to use Word and similar kinds of editing documents a lot more than something like LaTeX.

(Refer Slide Time: 11:59)

### Semantic Markup

- Content vs Presentation
- *Semantics*
  - Meaning of the text
  - structure or logic of the document



So, ultimately, in other words, what we are seeing here is, there is a distinction between content, that is the meaning of the content versus the presentation, how it is displayed? And the content is dealing more with what is called semantics. So, the semantics is something which once again keeps popping up every now and then in different contexts. Ultimately, what it means is just

what is the meaning of something, how should something be interpreted? So, what is the meaning of the text or what is the structure or logic behind this document?

And if we can put semantic markup into a set of documents, in general, that is more useful, because it means that this can be interpreted in different contexts appropriately. I still know what is the title of the document, which is the abstract, which is the main sort of the takeaway, what are the key words. If I have emphasized something, I will use emphasis rather than just bold font. And how you want to show emphasis can be decided by you. You want it to be in a different color, not just in a bold font. All of those are choices that you can make once you understand what the person was trying to convey.

