ASTER
Towards Display-Independent Electronic Information

T. V. Raman
Cambridge Research Lab
Digital Equipment Corporation

Outline

- Significance of electronic information.
- Overview of ASTER
- Extracting information structure.
- Audio Formatting with AFL.
- Obtaining multiple aural views.
- Browsing information structure.

$\boxed{\textbf{Significance}}$

Electronic Information
Is Display-Independent.

Can be accessed

- Where the user wants
- When the user wants
- How the user wants

Ubiquitous Multimodal Access.

$\boxed{\textbf{What is formatting?}}$

Information has structure:

- Bulleted lists.
- Mathematical expressions.

$$e^{e^{e^{x}}} e^{e^{x}} e^{x}$$

- Tables.

Formatted output:

- Permits direct perception of structure.
- Reduces cognitive load.

$\boxed{\textbf{Example}}$

A Math Expression

a, plus the fraction whose numerator is the expression, b plus c, end numerator, and whose denominator is the expression, d plus e, end denominator, plus x.	

$$\left\| \qquad\qquad\qquad \middle| \; a + \frac{b+c}{d+e} + x \; \right\|$$

Succinct notation essential to convey complex concepts.

The next-generation reading machine.

Information is display independent:

Electronic Document $\xrightarrow{\text{(LA)TEX}}$ Print

$\rightarrow$ 

Audio

Different views of the *same* document.

$\boxed{\textbf{Architecture}}$

$$\boxed{\textbf{Electronic documents}}$$

$$\Downarrow \text{Recognizer} \Downarrow$$

$$\boxed{\textbf{High-level model}}$$

$$\Downarrow \text{Audio Formatting} \Downarrow$$

$$\boxed{\textbf{Audio Rendering}}$$

$$+$$

$$\boxed{\text{Rendering Rules}}$$

$$+$$

$$\boxed{\text{Browsing}}$$

$$=$$

$$\boxed{\textbf{Audio Documents}}$$

Electronic Information

Documents are not just for printing!

Electronic	Printed
Logical Structure	Layout Structure
Dynamic	Static
Customizable	Fixed
Interactive	Non-Interactive
Active	Passive

Electronic Books

Make Information Come Alive!

Extracting Logical Structure

High-level models:

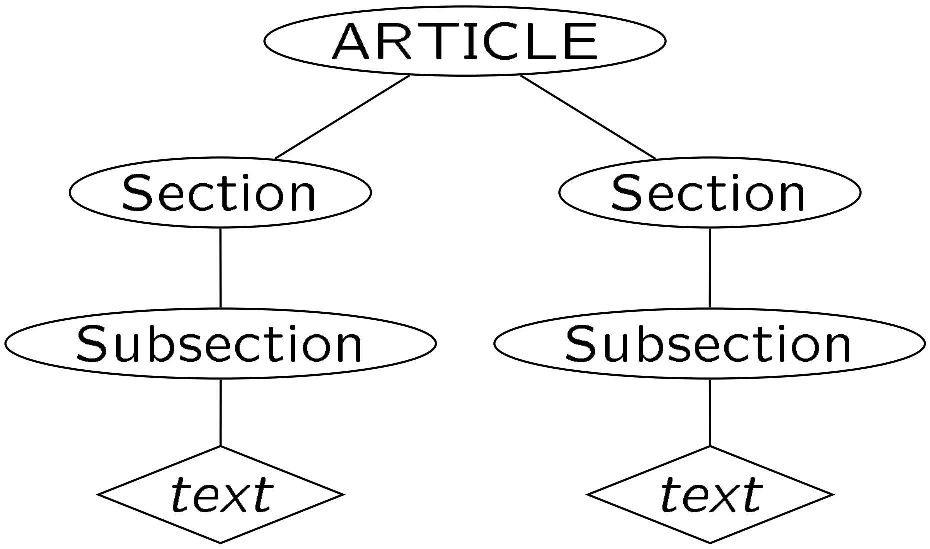
- **Tree structure** for text.
- **Quasi-prefix form** for mathematics.

A node in the quasi-prefix form:

$$\boxed{\begin{array}{ccc} \text{left-sup} & \text{accent} & \text{superscript} \\ \nwarrow & \uparrow & \nearrow \\ & \textbf{math object} & \\ \swarrow & \downarrow & \searrow \\ \text{left-sub} & \text{underbar} & \text{subscript} \end{array}}$$

Captures ambiguous written mathematics.

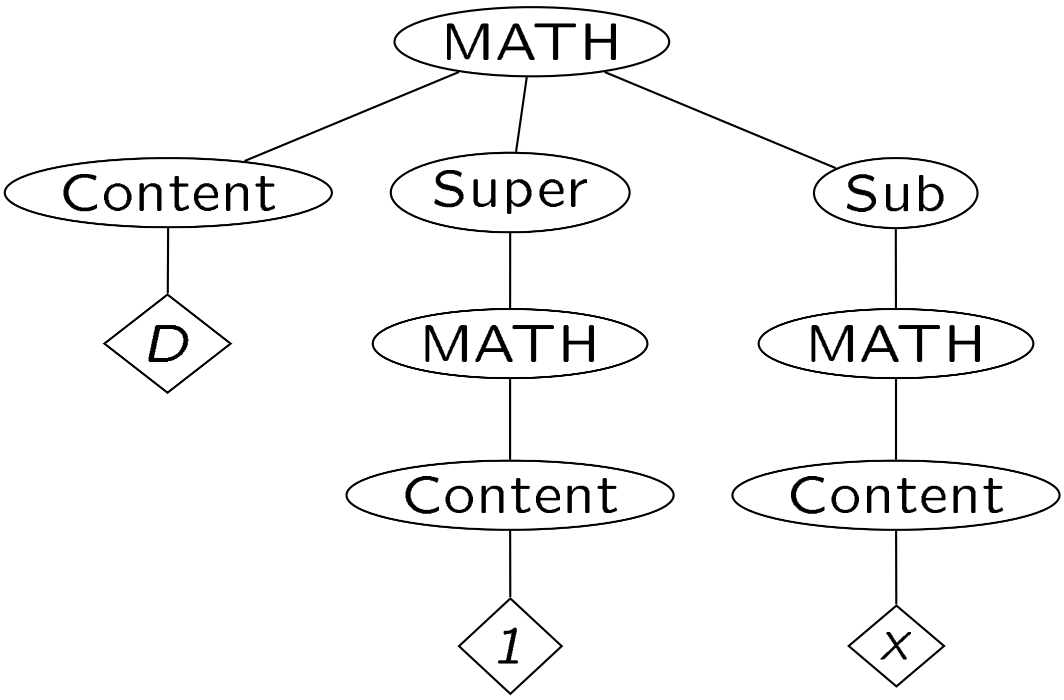
Example: Article



Example: Quasi-prefix form

$$D_x^1$$

is represented as :



Extending the recognizer

(LA)TEX macros introduce user-defined objects.

```
\newcommand\asterlogo{...}
```

Definition specifies rendering rule:

Object *asterlogo* can be rendered In Multiple ways.

$\boxed{\textbf{Defining the audio space}}$

Audio renderings: speech $\oplus$ sound.

Formatter is a Conceptual audio device:

- Formatter parameters define audio space.
- Formatter state is a point in this space.
- Formatter has multiple components.
- Components correspond to subspaces.
- Space = direct sum of subspaces.

Changing formatter state $\leftrightarrow$ Moving in audio space.

Audio Formatting Language

AFL is a block structured language.

AFL variables encapsulate formatter state.

Operators generate points in audio space.

Statements:

- **NEW-BLOCK** Introduce a local scope.
- **LOCAL-SET-STATE** Assignment.
- **GLOBAL-SET-STATE** Assignment.

Component spaces

Dimensions in the speech space:
Average-pitch, smoothness, directionality, ...

Operators: move-by, move-to, step-by, ...

Dimensions in the sound space:
volume, sound, ...

Operators: select-sound, switch-on, ...

$\boxed{\textbf{Producing audio renderings}}$

$\boxed{\text{Recognizer}} \xrightarrow{\textbf{AFL Rules}} \boxed{\text{Audio Renderings}}$

A typical rendering rule:

```
BEGIN-AFL-BLOCK
        set local state;
        render object
END-AFL-BLOCK
```

Nested objects rendered relative to parent.

$\boxed{\text{Mathematical notation}}$

Written notation:

- Linearizes implicit tree structure.

$$x + y + z$$

- Annotates this with attributes.

$$(x^2 + y^2)^2$$

A parallel audio notation:

Independent audio dimensions cue features.

Pick a monotonic change of state function f to vary audio state P along these dimensions, where the ordering $P < f(P) < f^2(P) < \cdots$ is perceptible.

This is sufficient to cue nested structures.

$\boxed{\text{Tree structure}}$

Move along a hyperplane in speech space spanned by six dimensions.

$$\frac{a+b}{c+d}$$

$$\frac{a}{b+c+d}$$

$$a+\frac{b+c}{d+e}+x$$

Can cue deeply nested structures.

$$1+\frac{x}{1+\frac{x}{1+\frac{x}{1+\frac{x}{1+\frac{x}{1+\ddots}}}}}$$

Superscripts and subscripts

$$x^{k_1} + x^{k_2} + x^{k_3} + \cdots + x^{k_n} = 0$$

$$x_{k^1} + x_{k^2} + x_{k^3} + \cdots + x_{k^n} = 0$$

$$e^{e^{e^x}} e^{e^x} e^x$$

$\boxed{\textbf{Complex notation}}$

Features can be combined:
Tree-structure within superscript.

$$x + y^{\frac{2}{k+1}}$$

Superscript within tree-structure.

$$\sqrt{e^{e^{e^x}} + \sqrt{e^{e^x} + \sqrt{e^x}}}$$

This system of audio notation scales very well.

Stereo Effects

Directionality is an audio dimension.
Moving along it produces stereo.

Let L be the line spanned by the dimensions left-volume and right-volume.

```
compute-step-size(length(row),L);
BEGIN-AFL-BLOCK
        set speech state  to left speaker;
        LOOP for  element in row do
                render  element ;
                step by 1 along L
        ENDLOOP
END-AFL-BLOCK
```

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Classification of audio cues

Speech and non-speech audio cues.

Also classified as:

- **Fleeting cue** Does not last.
- **Persistent cue** Persists across time.

Fleeting cues identify objects.

Persistent cues convey environment.

Rendering Styles

Alpha-equality commutes with substitution. If $t =_\alpha t'$ and $s =_\alpha s'$, then

$$t[s/x] \underset{\alpha}{=} t'[s'/x]$$

.

- Active rule overrides active style.

- Activate rules to customize styles.

Enables multiple views of a document.

Conclusion

- Archive high-level document source.
- Online interactive documents are multimodal.
- Networking enables ubiquitous access.

$\boxed{\textbf{Contributions}}$

- Design of speech-centric interfaces.
- Render information structure orally.
- Enable active listening.
- Help reduce listener's cognitive load.