EMMA

Extensible Multimodal Annotation markup language

Canonical structure semantic interpretations for a variety of inputs including:

- Speech
- Natural language text
- GUI
- Ink

EMMA

Extensible Multimodal Annotation markup language

Canonical structure semantic interpretations for a variety of inputs including:

- Speech
- Natural language text
- GUI
- Ink

Ink W3C standard: http://www.w3.org/2002/mmi/

EMMA

Represents user input

Vehicle for transmitting user's intention throughout application

Three components

- Data model
- Interpretation
- Annotation (main focus of standard)

General Annotations

Confidence

Timestamps

Alternative interpretations

Language

Medium (visual, acoustic, tactile)

Modality (voice, keys, photograph)

Function (dialog, recording, verification…)

EMMA Example

EMMA document

"I want to go from Boston to Denver on March 11, 2003"

```
<emma:emma emma:version="1.0" xmlns:emma="http://www.w3.org/2003/04/emma#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<emma:interpretation emma:id="int1">
                <origin>Boston</origin>
                <destination>Denver</destination>
                <date>03112003</date>
</emma:interpretation>
    <rdf:RDF>
            <!-- time stamp for result -->
            <rdf:Description rdf:about="#int1"
                    <emma:absolute-timestamp
                            emma:start="2003-03-26T0:00:00.15"
                            emma:end="2003-03-26T0:00:00.2"/>
            <!-- confidence score -->
            <rdf:Description rdf:about="#int1" emma:confidence="0.75"/>

        •<rdf:Description rdf:about="#int1"
        emma:model="http://myserver/models/city.xml"/>

    </rdf:RDF>

</emma:emma>
```

Interpretation

Annotations

Data Model

The *same meaning* with speech and mouse input

```
<emma:interpretation medium="acoustic" mode="voice"
id="int1">
    <origin>Boston</origin>
    <destination>Denver</destination>
    <date>03112008</date>
</emma:interpretation>
```

Speech

```
<emma:interpretation medium="tactile" mode="gui"
id="int1">
    <origin>Boston</origin>
    <destination>Denver</destination>
    <date>03112008</date>
</emma:interpretation>
```

Mouse

EMMA Annotations

- Tokens of input: emma:tokens attribute
- Reference to processing: emma:process attribute
- Lack of input: emma:no-input attribute
- Uninterpreted input: emma:uninterpreted attribute
- Human language of input: emma:lang attribute
- Reference to signal: emma:signal and emma:signal-size attributes
- Media type: emma:media-type attribute
- Confidence scores: emma:confidence attribute
- Input source: emma:source attribute
- Absolute timestamps: emma:start, emma:end attributes
- Relative timestamps: emma:time-ref-uri, emma:time-ref-anchor-point, emma:offset-to-start attributes
- Duration of input: emma:duration attribute
- Composite Input and Relative Timestamps
- Medium, mode, and function of user inputs: emma:medium, emma:mode, emma:function, emma:verbal attributes
- Composite multimodality: emma:hook attribute
- Cost: emma:cost attribute
- Endpoint properties: emma:endpoint-role, emma:endpoint-address, emma:port-type, emma:port-num, emma:message-id, emma:service-name, emma:endpoint-pair-ref, emma:endpoint-info-ref attributes
- Reference to emma:grammar element: emma:grammar-ref attribute
- Dialog turns: emma:dialog-turn attribute

Verification

Claiming to be 'charles foster kane', the user said 'rosebud', and the speaker verification engine accepted the claim with a confidence of 0.95.

```
<emma:emma version="1.0"
xmlns:emma="http://www.w3.org/2003/04/emma/">

  <emma:interpretation id="interp1 emma:duration="1810“
     emma:confidence="0.95" emma:process=file://myverifier
     emma:signal="http://example.com/signals/sg23.bin"
     emma:medium="acoustic“ emma:verbal="true"
     emma:mode="speech" emma:start="1149773124516"
     emma:uninterpreted="false" emma:function="verification"
     emma:dialog-turn="1“ emma:end="1149773126326"
     emma:lang="en-US" emma:tokens="rosebud" >

       <claim>charles foster kane</claim>

       <result>verified</result>

  </emma:interpretation>

</emma:emma>
```

If no ASR results are available, 'emma:tokens="rosebud"' would be omitted.

Identification

The user said 'rosebud' and the speaker identification engine identified the speaker as 'charles foster kane' with a confidence of 0.95.

```
<emma:emma version="1.0“
xmlns:emma="http://www.w3.org/2003/04/emma/">

  <emma:interpretation id="interp1" emma:duration="1810“
    emma:confidence="0.95" emma:process=file://myidentifier
    emma:signal=http://example.com/signals/sg23.bin
    emma:medium="acoustic“ emma:verbal="true"
    emma:mode="speech" emma:start="1149773124516“
    emma:uninterpreted="false" emma:function="identification"

    emma:dialog-turn="1" emma:end="1149773126326"
    emma:lang="en-US“ emma:tokens="rosebud" >

          <result>charles foster kane</result>

  </emma:interpretation>

</emma:emma>
```

Emma: fusion

Multiple sources of input

- Voice into a speaker verification engine
- Dialog into a VoiceXML 2.x engine

Results of both engines are represented using EMMA

Merging engine combines these two results into a single result

The three engines may be

- Co-located at a single site or distributed across a network
- May be performed in real time or delayed time

EMMA: fusion

Speech

Keyboard

Voice Samples

Speaker Identification

VoiceXML Engine

VoiceXML Dialog

EMMA

EMMA

Merging/ Unification

EMMA

Applications

EMMA: fusion

Speech

Keyboard

Grammar + Semantic Interpretation Instructions

Speech Recognition

Keyboard Interpretation

Interpretation Instructions

EMMA

EMMA

```
<interpretation mode = "voice">
  <emma:interpretation
id="interp1"
    emma:function="verification"
    emma:confidence="0.6">
     <result> John Dow </result>
</interpretation>
```

Merging/ Unification

EMMA

Applications

EMMA: fusion

Speech

Keyboard

Grammar + Semantic Interpretation Instructions

Speech Recognition

Keyboard Interpretation

Interpretation Instructions

EMMA

EMMA

```
<interpretation mode = "voice">
  <emma:interpretation
id="interp1"
    emma:function="verification"
    emma:confidence="0.6">
     <result> John Dow </result>
</interpretation>
```

Merging/ Unification

EMMA

Applications

```
<interpretation mode = "text">
  emma:interpretation id="interp1"
    emma:function="dialog"
    emma:confidence="0.6">
     <result> John Dow </result>
</interpretation>
```

EMMA: fusion

<interpretation mode = "derived">
emma:interpretation id="interp3"
emma:function="fusion"
emma:confidence="0.7">
<result> John Dow </result>
</interpretation>

Speech

Grammar + Semantic Interpretation Instructions

Speech Recognition

Interpretation Instructions

EMMA

<interpretation mode = "voice">
<emma:interpretation
id="interp1"
emma:function="identification"
emma:confidence="0.6">
<result> John Dow </result>
</interpretation>

Merging/ Unification

EMMA

Applications

<interpretation mode = "text">
emma:interpretation id="interp2"
emma:function="dialog"
emma:confidence="0.6">
<result> John Dow </result>
</interpretation>

Summary

EMMA can be used for many types of data

EMMA captures information about each data type

EMMA information is used in various processing phases

- Interpretation and semantic processing
- Fusion
- Data transmission