



Extensible Multimodal Annotation (EMMA) for Intelligent Virtual Agents

Deborah A. Dahl

Conversational Technologies

dahl@conversational-technologies.com

www.conversational-technologies.com

EMMA

EMMA 1.0 is a World Wide Web Consortium (W3C) XML standard for representing multimodal user inputs. EMMA is useful for supporting interoperability among the many components that can comprise a virtual agent, because it provides a very consistent way to represent the semantics and annotations for various inputs, including different user input modalities such as speech and handwriting as well as biometrics and sensor data.


EMMA for Different Modalities

Intent from speech, mouse, ink...

```

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

```



```

<emma:interpretation medium="tactile" mode="ink" id="int1">
  <origin>Boston</origin>
  <destination>Denver</destination>
</emma:interpretation>

```

Different annotations

Same semantics

Boston to Denver

Biometric identification from face, voice, fingerprint...

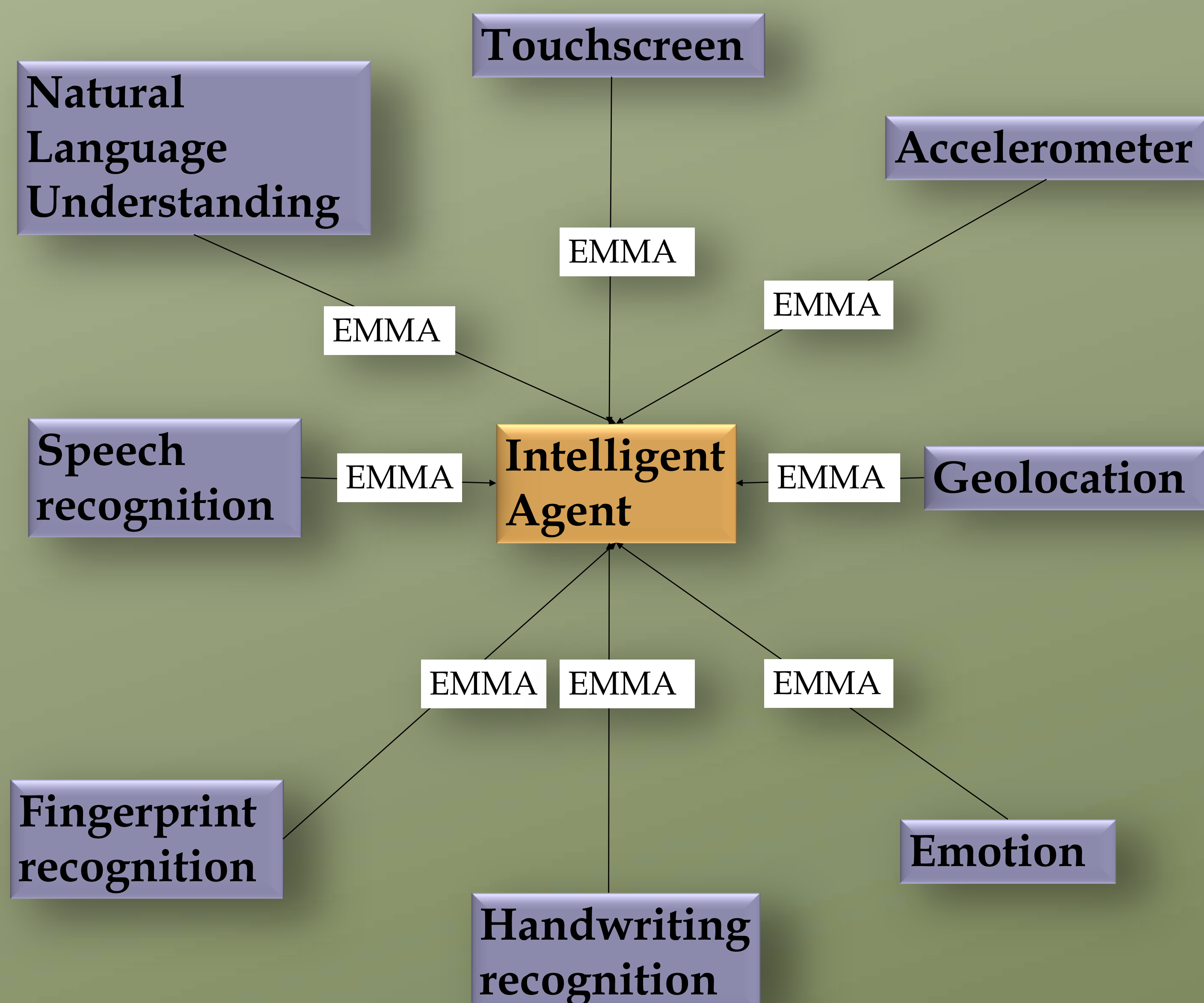
```

<emma:emma version="1.0">
  <emma:interpretation id="int1"
    emma:confidence="0.75"
    emma:medium="visual"
    emma:mode="photograph"
    emma:verbal="false"
    emma:function="identification">
    <person>12345</person>
    <name>Mary Smith</name>
  </emma:interpretation>
</emma:emma>

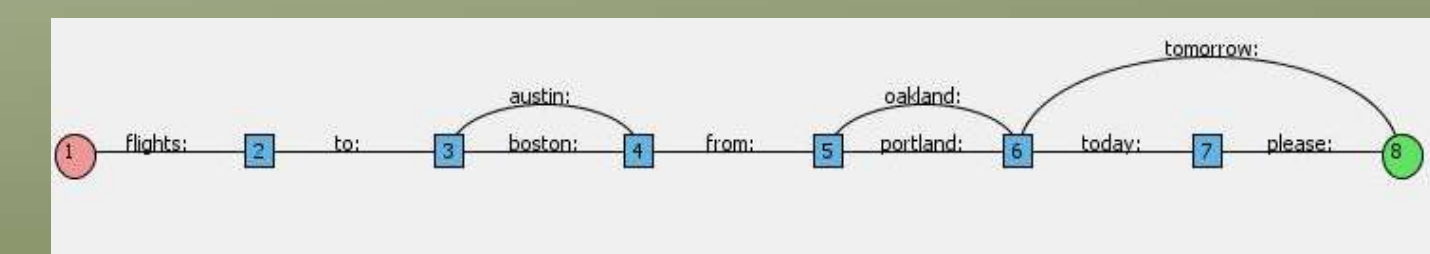
```



EMMA for Virtual Agents



Lattices for Uncertainty



Annotations

- Interpretation (Application Semantics)
- Tokens
- Confidence
- Timestamps (absolute and relative)
- duration
- Processor
- Medium
- Mode
- Function
- Verbal
- Alternatives (lattice or nbest)
- Signal
- Language
- Grammar
- Data model
- Derivation
- Groupings
- Sequences
- Source
- Extensibility - info
- Failure to interpret or no input
- Dialog turn

Multimodal Fusion of Speech and Ink with EMMA "Zoom in here"

Speech

```

<emma:emma version="1.0">
  <emma:interpretation id="voice1"
    emma:medium="acoustic"
    emma:mode="voice">
    <command>
      <action>zoom</action>
      <location emma:hook="ink">
        <type>area</type>
      </location>
    </command>
  </emma:interpretation>
</emma:emma>

```

Ink

```

<emma:emma version="1.0">
  <emma:interpretation id="pen1"
    emma:medium="tactile"
    emma:mode="ink">
    <location>
      <type>area</type>
      <points>42.13 -37.12 42.13 -37.12 ... </points>
    </location>
  </emma:interpretation>
</emma:emma>

```

Fused Input

```

<emma:emma version="1.0">
  <emma:interpretation
    emma:medium="acoustic tactile"
    emma:mode="voice ink"
    emma:process="http://example.com/myintegrator.xml">
    <emma:derived-from
      resource="http://example.com/voice1.emma/#voice1"
      composite="true"/>
    <emma:derived-from
      resource="http://example.com/pen1.emma/#pen1"
      composite="true"/>
    <command>
      <action>zoom</action>
      <location>
        <type>area</type>
        <points>42.13 -37.12 42.13 -37.12 ..</points>
      </location>
    </command>
  </emma:interpretation>
</emma:emma>

```

More Information

The specification: www.w3.org/TR/emma/
 Use Cases for EMMA 2.0: www.w3.org/TR/emma-usecases/
 The Multimodal Interaction Working Group: www.w3.org/2002/mmi/
 An open source implementation: <http://sourceforge.net/projects/nlworkbench/>
 Demos and examples: www.conversational-technologies.com/emma