Automated Retrieval of Technical Data from Scanned Drawing and PDF Images

DMSMS 2018
Dr. L. Robert Pokorny
XSB, Inc.
December 5, 2018

Work presented in this presentation funded in part through
DLA Contract SP4701-14-D-7018 Short Term project 8-X-02
Supporting R&D for Castings and Forgings

Distribution A: Approved for Public Release
Agenda

Background
- OCR processing Scanned Images of Part Drawings
- Extracting and Inferring Non-Geometric Part Properties
- Automating the Process
- Questions
Understanding Properties of Parts

• When the availability of a part is affected by DMSMS issues, understanding the characteristics of that part is the first step in finding a DMSMS solution.

• XSB, Inc. has worked with the Defense Logistics Agency over the last 20 years to aggregate critical properties of National Stock Numbers (NSNs,) from DOD and external data:
  – Specifications
  – Material
  – Manufacturing Processes
  – Size
  – Usage

• We use Artificial Intelligence technology to extract and infer critical properties of NSNs from unstructured data.

• Aggregated critical NSN data is maintained in the Coherent View® database that we maintain and update quarterly under contract with the Defense Standardization Program Office. The Coherent View supports:
  – Pin Point® part research web tool
  – Weapon System Impact Tool (WSIT)
What Coherent View Does

- What happens when important data is missing?
- XSB mines enterprise and external data to infer and enrich properties for the item.

Superseded by ASTM-B26/B26M

Infer Manufacturing Process:
Sand Casting
Material: Aluminum Alloy 356.0
Temper: T6

MECHANICAL PROPERTIES
Tensile Strength: 30,000 psi
Yield Strength: 20,000 psi
Extension under Load: .0039 in/in

<table>
<thead>
<tr>
<th>CONSTITUENT ELEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
</tbody>
</table>
Agenda

• Background

OCR processing Scanned Images of Part Drawings

• Extracting and Inferring Non-Geometric Part Properties

• Automating the Process

• Questions
Aggregated Properties Only Go So Far

- Coherent View® aggregates critical NSN properties from Federal Logistics Information System, (FLIS,) technical characteristics and DLA Enterprise Business System, (EBS,) logistics data
- Pin Point can narrow down to parts of interest, but final decisions often require manually examining the Technical Data Package, (TDP,) for an NSN
  - Coherent View® indicates when an NSN is a possible or likely casting or forging
  - Conclusive evidence for this often can only be discovered in drawing notes from C4 drawing or PDF image files in the TDP
- TDPs are only available for an NSN on an as needed basis so critical data in them can not be pre-aggregated
- What is needed is a process to apply AI extraction and inference technology to TDP drawings on demand
  - Optical Character Recognition, (OCR,) technology to convert non-geometric drawing data to text
  - XSB AI technology to extract and infer critical part properties from text
GENERAL NOTES - RAW FORGING

51: SPECIFIED TOLERANCES INCLUDE DIE CLOSURE, LINEAR, STRAIGHTNESS AND MISMATCH TOLERANCES AS APPLICABLE.

52: DRAFT ANGLES 5° ± .03 MATCHED WHERE NECESSARY.

53: CORNER RADIUS .16 ± .03 EXCEPT AS NOTED.

54: FILLET RADIUS 0.12 ± 0.03 EXCEPT AS NOTED.

55: PADS SHALL BE IN AS-FORGED CONDITION, NO GRINDING PERMITTED.

56: MAXIMUM FLASH EXTENSION .03.

57: RECORDS OF MILL HEAT AND HEAT TREATMENT TO BE FURNISHED TO VERTOL.

58: MARK PER VERTOL SPEC. MS 1301 GRADE IB.

59: FABRICATE FORGING IN ACCORDANCE WITH SPEC. QQ-M-40.

60: DATUM DIMENSION (ZERO TOLERANCE) FOR LOCATION OF DATUM PLANE.

ALTERNATE MATERIAL - ZK60A-T6 MAGNESIUM ALLOY PER QQ-M-31 OR AMS 4352. STOCK SIZE 3.00 X 5.80 X 5.10
GENERAL NOTES - RAW FORGING
‘S1: SPECIFIED TOLERANCES INCLUDE: DIE CLOSURE, LINEAR, STRAIGHTNESS AND MISMATCH TOLERANCES AS APPLICABLE
‘S2: DRAFT ANGLES 5° MATCHED WHERE NECESSARY.
‘S3: CORNER RAMI J6 2.03 EXCEPT AS NOTED.
‘S4: FILLET RADII .12 203 EXCEPT AS NOTED.
[Be>rans ‘SHALL BE IN AS-FORGED CONDITION NO GRINDING PERMITTED.
‘S6: MAXIMUM FLASH EX TENSION .03.
[BE>RECORDS OF MILL HEAT AND HEAT TREATMENT TO BE FURNISHED TO VERTOL
D [BB MARK PER VERTOL SPEC. Ms fame GaTAOn TE
59> FABRICATE. FORGING IN ACCORDANCE, WITH SPEC. QQ-M-40.
[> datum omissions (ZERO TOLERANCE) FOR LOCATION OF DATUM PLANE.
> ALTERNATE MATERIAL - ZK60A-TS MAGNESIUM ALLOY PER QQ-M-31 OR AMS 4352. STOCK SIZE 3.00 X 5.80 X 5.10
Agenda

• Background
• OCR processing Scanned Images of Part Drawings
• Extracting and Inferring Non-Geometric Part Properties
• Automating the Process
• Questions
Specification References in Technical Data

- TDPs often refer to specifications for requirements about materials or manufacturing processes
- Critical properties of NSNs can be inferred from these references
- XSB’s Semantic Web for Interoperable Specifications and Standards (SWISS) creates digital models of specifications from which part properties can be derived
- TDP references to specifications processed by AI technology can link directly to the specification in SWISS Spectacle®
Properties Derived from a Chain of References

Typical multi-document reference pattern

- **NSN**
  - **Has**
  - **Governed by**

- **Property**
- **As Per**
- **Specification**

**Drawing**

Pattern infers Secondary property

**Specification**
Properties Derived from a Chain of References

Typical Multi-spec reference pattern

- **NSN**
  - Is governed by
  - **QQ-M-40**
  - In Accordance with
  - **Drawing**
  - Forging

Pattern infers material can only be one of 7 Magnesium Alloys

Link to SWISS
FEDERAL SPECIFICATION
MAGNESIUM ALLOY FORGINGS

This specification was approved by the Commissioner, Federal Supply Service, General Services Administration, for the use of all Federal agencies.

1. SCOPE AND CLASSIFICATION

1.1 Scope. This specification covers the requirements of fully worked magnesium alloy forgings.

1.2 Classification.

1.2.1 Compositions. Magnesium alloy forgings covered by this specification shall be of the following compositions as specified (see 6.2):

- AZ31B—Magnesium—3.0 Aluminum—1.0 Zinc
- AZ61A—Magnesium—6.5 Aluminum—1.0 Zinc
- AZ60A—Magnesium—6.5 Aluminum—0.5 Zinc
- HM21A—Magnesium—2.0 Thorium—0.5 Manganese
- MA—Magnesium—1.2 Manganese
- TA54A—Magnesium—3.5 Aluminum—5.0 Tin
- ZK60A—Magnesium—5.5 Zinc—0.45 Zirconium
Agenda

• Background
• OCR processing Scanned Images of Part Drawings
• Extracting and Inferring Non-Geometric Part Properties
• Automating the Process
• Questions
Applying OCR and Extraction to TDPs
Automating the Process

TDP Tool

TDP

OCR processing

XSB extraction

Returned Evidence for Casting or Forging
## Job Summary

### Filter
Enter at least 5 characters

<table>
<thead>
<tr>
<th>ID</th>
<th>Source File</th>
<th>Submission Time</th>
<th>Status</th>
<th>Casting Possibility</th>
<th>Forging Possibility</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>171</td>
<td>4069T04_99207_DD_A_0001.C4</td>
<td>11/12/18, 1:33 PM</td>
<td>✔</td>
<td>high</td>
<td></td>
<td>SAE-AMS645 (UNCATEGORIZED), MIL-STD-8879 (SPECIFICATION), THREADING (PROCESS), STEEL ALLOY UNS S32100 (MATERIAL), POSSIBLE FORGING (PROCESS)</td>
</tr>
<tr>
<td>170</td>
<td>5044T93_99207_DD_D_0001.C4</td>
<td>11/12/18, 1:33 PM</td>
<td>✔</td>
<td>high</td>
<td>medium</td>
<td>ASTM-E122 (SPECIFICATION), SAE-AMS5832 (UNCATEGORIZED), SAE-AMS5552 (UNCATEGORIZED), RADIOLOGIC INSPECTION (PROCESS), ALUMINUM (MATERIAL), STEEL ALLOY UNS S21904 (MATERIAL), QUALITY PROCESS (PROCESS), CASTING</td>
</tr>
</tbody>
</table>
### Job 170

<table>
<thead>
<tr>
<th>Source File</th>
<th>Reference Type</th>
<th>Reference</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>MATERIAL</td>
<td>ALUMINUM</td>
<td>AI SHALL CONFORM TO</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>SPECIFICATION</td>
<td>ASTM-E122</td>
<td>ASTM E122</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>CASTING</td>
<td>4 UNLESS OTHERWISE SPECIFIED CASTING CONTOUR MUST NOT DEVIATE IN</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>CASTING</td>
<td>IN PROCESS WELDING OF CASTINGS)</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>CASTING</td>
<td>PRESSURE CASTING</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>FORGING</td>
<td>(CAST &amp; FORGED PARTS)</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>INVESTMENT CASTING</td>
<td>mat: AMS 5362 ENGRG ☻ PEARSON</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>QUALITY PROCESS</td>
<td>ASTM E122</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>PROCESS</td>
<td>RADIOLOGIC INSPECTION</td>
<td>RADIOGRAPHIC INSPECTION REQUIREMENTS</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5044T93_99207_DD_D_0</td>
<td>UNCATEGORIZED</td>
<td>SAE-AMS5362</td>
<td>mat: AMS 5362 ENGRG ☻ PEARSON</td>
</tr>
<tr>
<td>001.C4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Heuristics for probability assignment

CASTING is mentioned as a spec reference?

Yes

CASTING mentioned as a possible spec reference?

No

Count: CASTING mentioned in text independent from FORGING

Yes

Count > 3

No

E.g. “AMS 5382”

HIGH Probability

E.g. “Steel Castings, Sand and Centrifugal, Corrosion Resistant”

E.g. “MIL-C-19052”

MEDIUM

CASTING mentioned but always alongside FORGING

No

E.g. “Castings, forgings, extrusions and machined parts shall be…”

YES

LOW Probability

NO CASTING EVIDENCE FOUND
Questions