

DITA as a Building Block for Creating & Delivering Content in New Ways

Keith Schengili-Roberts • April 24, 2018





Agenda



1. Introduction
2. Building with DITA
 - Lightweight DITA in Software Development Workflow
 - iiRDS and Information 4.0
 - Chatbots and DITA
3. Conclusions
4. Q/A



Keith Schengili-
Roberts

Who's This Guy?



What I do:

- DITA Evangelist & Market Researcher at IXIASOFT
- Chair of OASIS DITA Adoption Committee
- Member of OASIS DITA Technical Committee and LwDITA Sub-committee
- Lecturer on Information Architecture at the University of Toronto
- 12+ years of experience with DITA XML



AKA “DITAWriter”

- Industry blog started 6+ years ago
- Over 297,000 hits(!)
- Regularly updated info on DITA conferences, DITA books, companies using DITA, DITA CMSes, DITA editors, other DITA tools, and DITA consulting firms
- News and views on DITA use
- Also features interviews with those making a difference in the world of DITA

The screenshot shows the homepage of the DITAWriter website. At the top, there is a navigation bar with links for ARTICLES, NEWS, INTERVIEWS, WEBINARS, SOFTWARE UPDATES, DITA CONSULTANTS, and ABOUT. A search bar is located on the right. Below the navigation bar is the site's logo, which includes the text "<di>a>Writer" and a circular icon with the word "STRATEGY". A secondary navigation bar lists categories: DITA CONFERENCES, DITA BOOKS, COMPANIES USING DITA, DITA CMSes, DITA EDITORS, and DITA TOOLS. The main content area features a large banner image with a clock face and the text "DON'T WAIT FOR THE DITA 2.0 STANDARD TO MIGRATE TO DITA". Below the banner are tabs for "Latest Blogs", "Popular", and "Recommended". The "Latest Blogs" section displays two articles: "SIGDOC 2017: EDUCATING TOMORROW'S TECHNICAL COMMUNICATORS" and "DON'T WAIT FOR THE DITA 2.0 STANDARD TO MIGRATE TO DITA". On the right side, there is a "FOLLOW ME ON TWITTER" section with a tweet from @ditawriter and a "RECENT POSTS" section listing various articles.



The Inspiration for this Presentation

- As a presenter, I attend 5-6 major industry-related conferences per year
- Last year, I was intrigued to find a definite shift in the talk about DITA, less as a means in itself, but as a building block for other technologies
 - None of these were DITA-specific conferences
- Customers have also expressed their own specific needs for how they intend to use DITA





Some Overlooked Advantages of DITA

- DITA was built *primarily* with reuse optimization in mind, but it also has these other cool features which make it ideal as a building block:
 - Structured, typed content that is also modular
 - Can work well with other formats
 - Short descriptions as a way of summarizing content for users to find appropriate content
- Along with content reuse, these features help make DITA a good building block for use with other technologies

A hand is shown in the upper right corner, holding a single wooden block. Below it, a stack of several wooden blocks is visible, arranged in a row. The background is a soft, out-of-focus green and white, suggesting an outdoor setting. The text is centered on an orange rectangular background that spans the width of the image.

Building with DITA: Lightweight DITA in Software Development Workflow



The Rise of Markdown in Programming Environments

- In recent years there has been a shift from using XML or HTML for programming documentation
- One example: Microsoft has moved the authoring of their developer/IT docs to Markdown
- Clients and industry contacts confirm that there is a call for being able to work with Markdown content with DITA

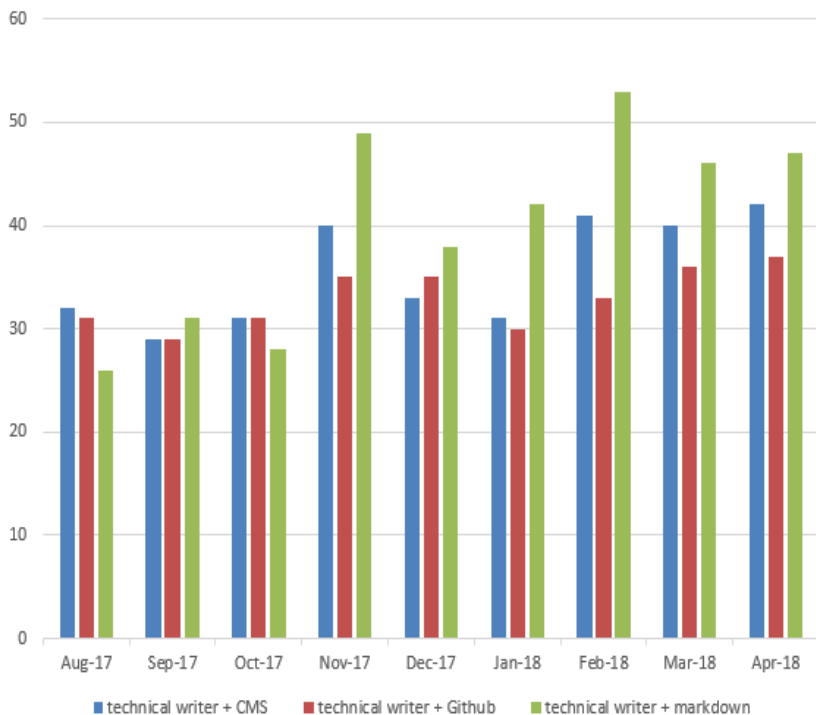
The screenshot shows two versions of a document. The top version is from ars TECHNICA, featuring a dark header with the site name, a 'SUBSCRIPTIONS' button, and a search icon. The article title is 'Microsoft unveils new effort to make its developer, IT documentation great again' by Peter Bright, dated 5/4/2016. The content discusses Microsoft Intune's security features. The bottom version is from TechNet, showing a similar article layout with a navigation menu and a search bar. The caption below the screenshot reads: 'Above: the new docs.microsoft.com appearance. Below: the same article in old TechNet.'

Microsoft's developer documentation used to be the model that all others should follow. The documentation itself was thorough, combining reference material with usage guides and sample code. Its use of, at the time, novel JavaScript and XML techniques (known in those days as dynamic HTML, or DHTML) made it easy to browse through the documentation and quickly switch between

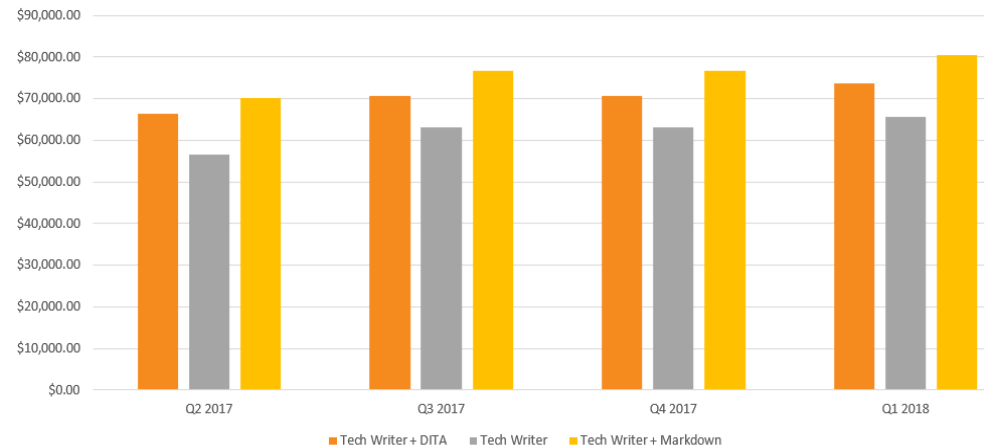


Markdown in Tech Docs More Prevalent

Rise of New Technologies/Tools in Technical Writer Job Postings on Indeed.com



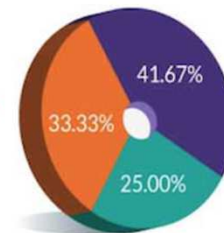
Avg. Wage Rates for Technical Writers + Specific Skill per Quarter from Q2 2014 - Q1 2018



Q2 What are the tools you use today to create content?

Markdown or similar

- Primary 41.67%
- Secondary 25.00%
- Tertiary 33.33%



From the CIDM/DCL 2018 Following the Trends survey

- But one thing Markdown doesn't do at all: reuse!



Lightweight DITA (MDITA) to the Rescue!

- This is one of the scenarios that led to the development of Markdown-flavoured LwDITA (MDITA)
- Official OASIS Committee Note outlining usage is now available
- MDITA comes in two types:
 - MDITA Core (GitHub-flavored, no keys)
 - MDITA Extended (GitHub with YAML headers and keys)



Workflows that MDITA Makes Possible

- Scenario #1: SME writes content in MDITA, Technical Writer can then edit content and import final results to “full” DITA 1.3
- Scenario #2: SME writes content in MDITA Core, Technical Writer edits content (in MDITA or XDITA), roundtrips changes back in MDITA back to SME using DITA-OT to convert content



The Future is Now

- Oracle is testing Markdown to DITA, and plans to trial MDITA later this year
- Other companies (primarily within the software sector) are also using Markdown and migrating the content to DITA





Want to Try it Yourself?

- I have created a set of LwDITA code samples in all of the major flavours, available from: [github.com/DITAWriter/LwDITA Code Samples](https://github.com/DITAWriter/LwDITA_Code_Samples)

DITAWriter / LwDITA_Code_Samples

Unwatch 2 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

A set of code samples designed to demonstrate what is possible for each of the "flavors" of Lightweight DITA. Edit

Add topics

36 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

Commit	Message	Time
DITAWriter	Add files via upload	Latest commit 45160bd 3 hours ago
TRIS-80 Expansion Module (HDITA - With Keys)	Add files via upload	3 hours ago
TRIS-80 Expansion Module (MDITA Core)	Add files via upload	a month ago
TRIS-80 Expansion Module (MDITA Expanded - No Keys)	Add files via upload	3 hours ago
TRIS-80 Expansion Module (MDITA Expanded - With Keys)	Add files via upload	3 hours ago
TRIS-80 Expansion Module (XDITA - With Keys)	Delete index.map	3 hours ago
README.md	Update README.md	7 days ago

README.md

LwDITA_Code_Samples

A set of code samples designed to demonstrate what is possible for each of the "flavors" of Lightweight DITA. All examples are based on sections of various TRS-80 manuals from the late-1970s/early-1980s that have either lapsed out of copyright or have been donated to the Public Domain. Includes sample marketing material, computer hardware content, software samples, legal frontmatter and glossary.

These files represent one interpretation as to how Lightweight DITA markup ought to be done, based on information available

A hand is shown in the upper right corner, holding a small, rectangular wooden block. The background is a soft, out-of-focus green field under a bright sky. At the bottom of the image, a row of several similar wooden blocks is laid out horizontally. The overall composition is clean and minimalist, suggesting a theme of building or construction.

Building with DITA: DITA Metadata and SEO



What Search Engines Do and Why

- Search engines want to provide users with the most relevant result to a given search query
- This is why product documentation is important from an SEO perspective, as users may find relevant answers to their queries from tech docs as opposed to a *poorly-focused* marketing communication that does not focus squarely on users' needs

High-quality technical information is a decisive factor for purchase decision

IBM has been developing a technical communications body of knowledge with a focus on content as a business asset. The company runs a permanent survey on its website asking customers to evaluate the importance of high-quality technical information. The responses of 179 participants of this survey between 2010 and 2013 have been summarized in an article for the Technical Communication Magazine of the STC.

Respondents overwhelmingly agreed that high quality technical information is important or very important with regard to the

- the perception of the company (84.5%)
- the satisfaction with the product or solution (93.3%)
- the perception of the product quality (96.4%) and
- the initial purchase decision (88.7%).

From "A business case for technical communication – facts & figures"
www.tcworld.info/e-magazine/technical-communication/article/a-business-case-for-technical-communication-facts-figures/



Tech Docs are Not Blogs

- A lot of SEO material out there is focused on ranking marketing material and blog content; there are only a handful of articles on tech docs and SEO
- Many typical SEO recommendations may not apply or are outside of your control
- Technical documentation should be factored into an overall corporate SEO strategy

Domain Factors



1. **Domain Age:** In [this video](#), Matt Cutts states that:

“The difference between a domain that’s six months old versus one year old is really not that big at all.”

In other words, they *do* use domain age...but it’s not very important.

2. **Keyword Appears in Top Level Domain:** Doesn’t give the boost that it used to, but having your keyword in the domain still acts as a relevancy signal. After all, they still bold keywords that appear in a domain name.

3. **Keyword As First Word in Domain:** A domain that starts with their target keyword has an edge over sites that either don’t have the keyword in their domain or have the keyword in the middle or end of their domain.

4. **Domain registration length:** A [Google patent](#) states:

“Valuable (legitimate) domains are often paid for several years in advance, while doorway (illegitimate) domains rarely are used for more than a year. Therefore, the date when a domain expires in the future can be used as a factor in predicting the legitimacy of a domain”.

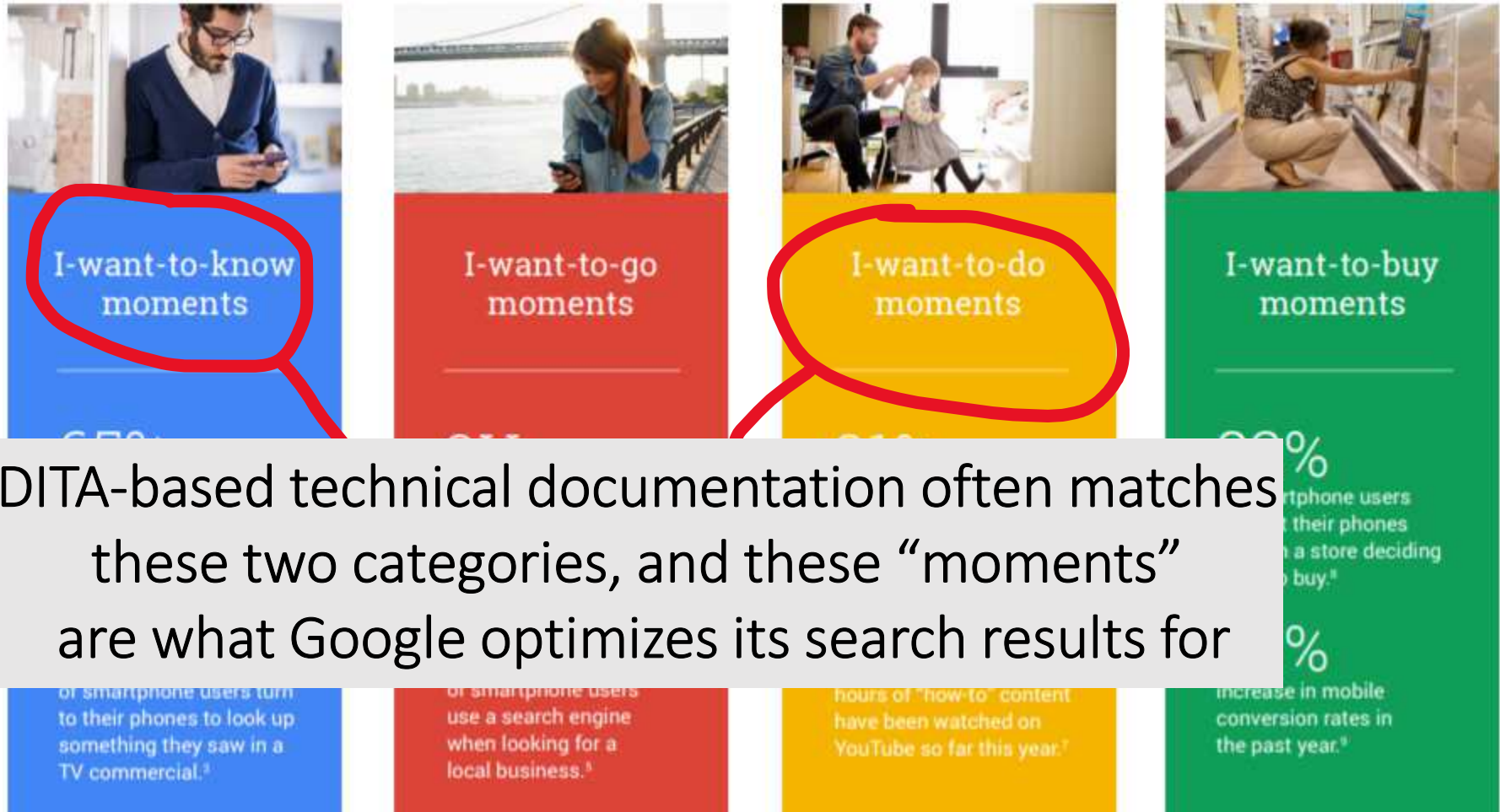
BackLinko-Google’s 200 Ranking Factors:
backlinko.com/google-ranking-factors

How DITA Content Can Aid with SEO

- I have given whole talks on this topic before. In a nutshell:
 - Use `shortdesc` because Google uses it in search results* and it provides info to users as to whether or not it contains the info they need
 - Have a naming convention for your DITA output that is descriptive:
`Understanding_the_Fliputron_feature_on_the_Vebulon_5.html`
 - Use descriptive `<title>`s that are no longer than 75 characters long
 - Relationship tables = internal links which increase content relevancy
 - And last, but far from least: content that targets what users are looking for!

* When written well

Also, Keep in Mind Google's Four User "Moments"



DITA-based technical documentation often matches these two categories, and these “moments” are what Google optimizes its search results for

I Want to Know / I Want to Do and DITA

“I Want to Know” moments match with DITA concept topics:

- “Concepts provide background that helps readers understand essential information about a product, a task, a process, or any other conceptual or descriptive information.”

“I Want to Do” moments match with DITA task topics:

- “A task information type.. provid[es] precise step-by-step instructions detailing the... actions that must be performed.”





I Want to Know / I Want to Do and DITA

- What can you do?
 - Write your titles and content specifically to answer potential questions that can be posed by your users
 - Creating user scenarios can help you map out likely questions and answers
 - Further emphasizes idea of “one goal per topic”



Use Schema.org to Help Identify Your Content

- Sponsored by Google, Bing, Yandex and Yahoo! to “create and support a common set of schemas for structured data markup on web pages.”
- Its vocabulary is designed for marking up content with semantic descriptions aimed at web spiders
- Uses Microdata, RDFa, or JSON-LD formats

Sample Schema.org Code Rendered in RDFa

```
<div vocab="http://schema.org/" typeof="Movie">
  <h1 property="name">Avatar</h1>
  <div property="director" typeof="Person">
    Director: <span property="name">James Cameron</span>
    (born <time property="birthDate" datetime="1954-08-16">August 16, 1954</time>)
  </div>
  <span property="genre">Science fiction</span>
  <a href="../movies/avatar-theatrical-trailer.html" property="trailer">Trailer</a>
</div>
```

How This Can Be Applied to Existing Technical Content

Google Structured Data Testing Tool



NEW TEST

```
1 <script type="application/ld+json">{
2   "@context": "http://schema.org",
3   "@type": "TechArticle",
4   "proficiencyLevel": "Beginner",
5   "datePublished": "March 27 2017",
6   "headline": "Insert a new image",
7   "image": {
8     "@type": "imageObject",
9     "url": "http://www.ixiasoft.com/files/5015/1395/7057/per1389986003493.png",
10    "height": "115",
11    "width": "351"
12  },
13  "wordCount": "965",
14  "articleBody": "You can import a new image into the Content Store and insert it
15  at the cursor position in your topic."
16  },
17  "publisher": {
18    "@type": "Organization",
19    "name": "IXIASOFT",
20    "logo": {
21      "@type": "imageObject",
22      "url":
23        "http://www.ixiasoft.com/files/cache/42cf035cb32d36d8564741a5098efee0_f3926.png"
24    }
25  }
26 }
27 }
28 }</script>
```

TechArticle All (1) ▾

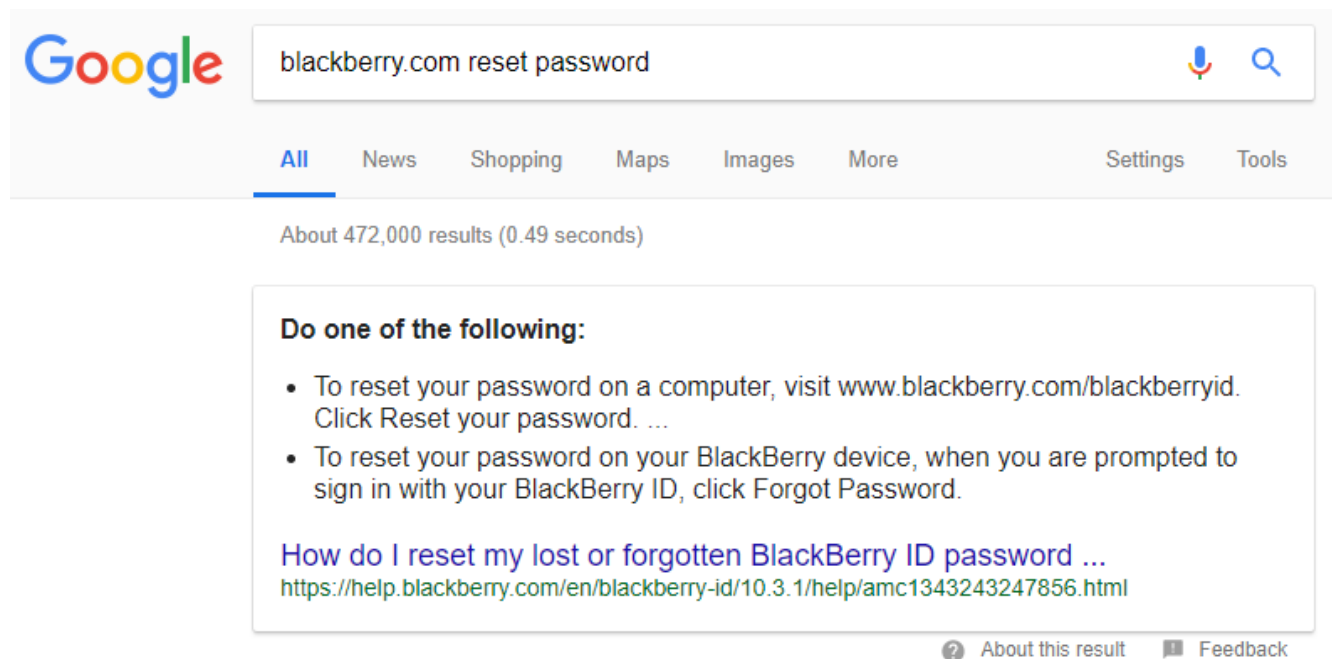
TechArticle 0 ERRORS 0 WARNINGS ^

@type	TechArticle
proficiencyLevel	Beginner
datePublished	2017-03-27
headline	Insert a new image
wordCount	965
articleBody	You can import a new image into the Content Store and insert it at the cursor position in your topic.
image	
@type	ImageObject
url	http://www.ixiasoft.com/files/5015/1395/7057/per1389986003493.png
height	
@type	Intangible
name	115
width	
@type	Intangible
name	351

<https://search.google.com/structured-data/testing-tool>

What Are the Benefits of This?

- Can *potentially* turn your content on Google into a “Featured Snippet” in searches.



The image shows a screenshot of a Google search interface. The search bar contains the text "blackberry.com reset password". Below the search bar, there are navigation tabs for "All", "News", "Shopping", "Maps", "Images", "More", "Settings", and "Tools". The "All" tab is selected. Below the tabs, it says "About 472,000 results (0.49 seconds)". A featured snippet box is displayed, containing the following text:

Do one of the following:

- To reset your password on a computer, visit www.blackberry.com/blackberryid. Click Reset your password. ...
- To reset your password on your BlackBerry device, when you are prompted to sign in with your BlackBerry ID, click Forgot Password.

Below the list, there is a link: [How do I reset my lost or forgotten BlackBerry ID password ...](https://help.blackberry.com/en/blackberry-id/10.3.1/help/amc1343243247856.html)
<https://help.blackberry.com/en/blackberry-id/10.3.1/help/amc1343243247856.html>

At the bottom right of the snippet box, there are two links: "About this result" and "Feedback".

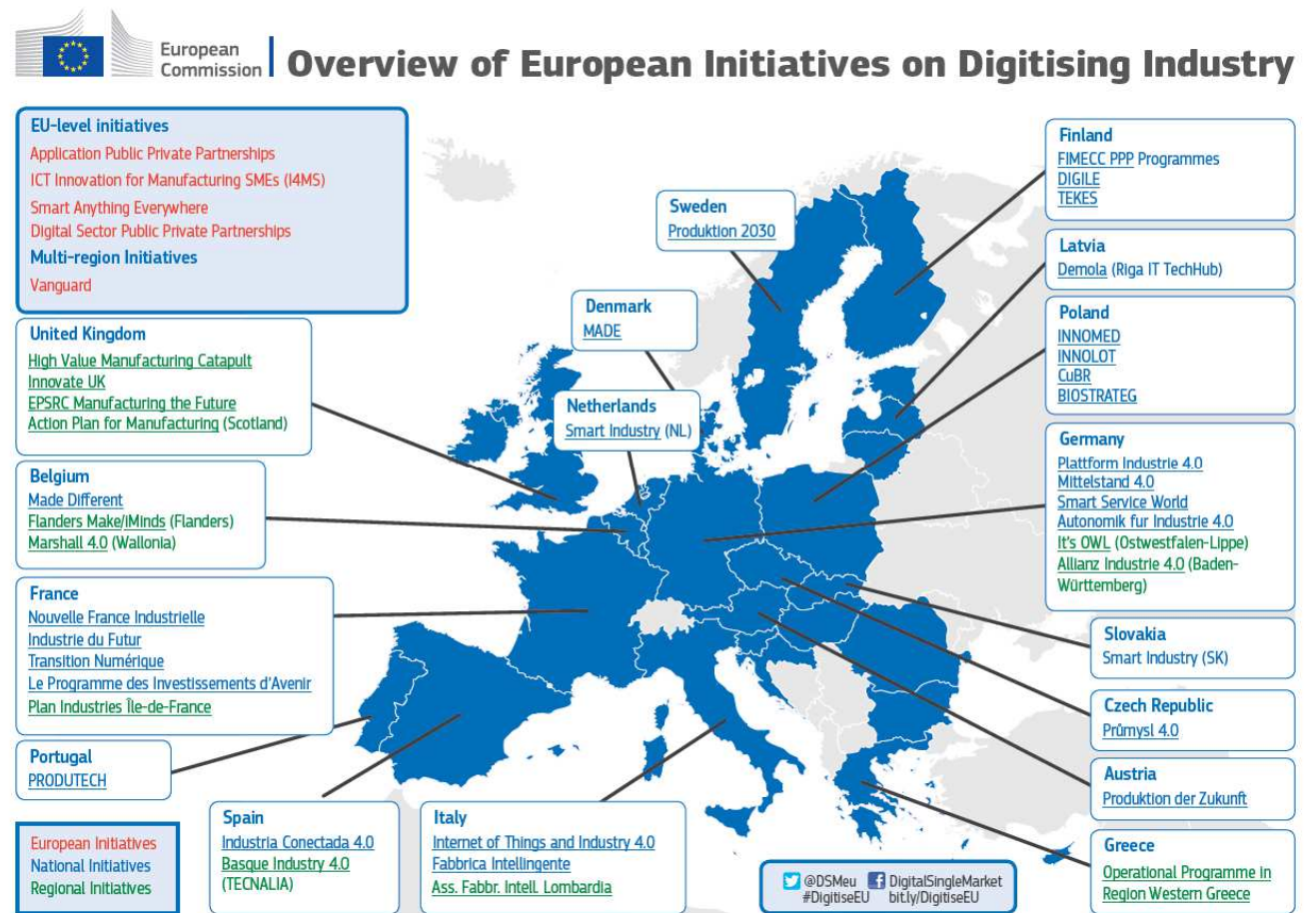
A hand is shown in the upper right corner, holding a single wooden block. Below it, a row of several wooden blocks is laid out horizontally. The background is a soft-focus green field under a bright sky. An orange banner is positioned in the middle of the image, containing the title text.

Building with DITA: iiRDS and Information 4.0



What is Information 4.0?

- Strong push for this within Europe, and while it may go under different names (as you see here in this EU infographic) in a nutshell it is about bringing together automation and data exchange within manufacturing systems





Best Single-slide Explanation of Information 4.0

Information 1.0	Information 2.0	Information 3.0	Information 4.0
One Format One Owner One Delivery One Publisher	Many Formats One Owner One Delivery One Publisher	Many Formats Many Owners Many Deliveries One Publisher	Many Formats Many Owners Many Deliveries Many Publishers
The Emergence of Content 4.0 for Industry 4.0			
Content 1.0 Publications	Content 3.0 Topics	Content 3.0 Components	Content 4.0 Molecules & Objects
 Illustration collaboratively produced by Marie Girard (IBM) & Joe Gollner (Gnostyx)			
Industry 1.0	Industry 2.0	Industry 3.0	Industry 4.0

- Industry 3.0 roughly represents where we are now
- Industry 4.0 is “smart content” where content is managed at a fine-grained level and incorporates information on how, when and where they behave

Image courtesy of Joe Gollner



Many Publishers, Collaborative Documentation

- Consider a device built from many parts from different manufacturers (such as a train diesel engine)
- Ideally, for maintenance purposes, it requires a manual that documents all of its component parts made by multiple manufacturers
- iiRDS can step in, creating: “dynamic documentation that can be assembled according to application and context”



Diesel train engine model from 3DHorse.com



What is iiRDS?

- iiRDS: “intelligent information Request and Delivery Standard”, designed to enable applications to exchange technical documentation across suppliers and devices
- Each real-world object is reflected in a digital asset *described* by iiRDS
- Firms create topics in bulk, and iiRDS describes these topics and their relationships, so that a document can be generated on demand based on context
- iiRDS metadata can also enrich content and enable semantic searches, search facets, etc.

tekom Request for Comments

iiRDS Specification

intelligent information Request and Delivery Standard



Request for Comments - 20 October 2017

Editors:

Achim Steinacker, [intelligent views](#)
Jan Oevermann, [ICMS](#)
Mark Schubert, [parson](#)
Markus Wiedenmaier, [practice innovation](#)
Martin Kreutzer, [Empolis](#)
Sebastian Göttel, [SCHEMA](#)
Ulrike Parson, [parson](#)
Win Nuding, [cognitas](#)

Author:

tekom Working Group Information 4.0, [List of contributors](#)

Implementation:

[RDF Schema](#)

Previous Version:

[First Public Working Draft](#)

Other Resources:

[Website](#)
[License](#)

© 2016-2017 Gesellschaft für Technische Kommunikation – tekom Deutschland e.V. All rights reserved.
This work is licensed under a [Creative Commons Attribution-NoDerivatives 4.0 International License](#).



https://iirds.tekom.de/fileadmin/tekom_iirds/iirds_specification/index.html



iiRDS Describes Relationships Between Objects

- iiRDS is based on an ontology that describes relationship between content objects, described as a metadata triplet, typically using RDF

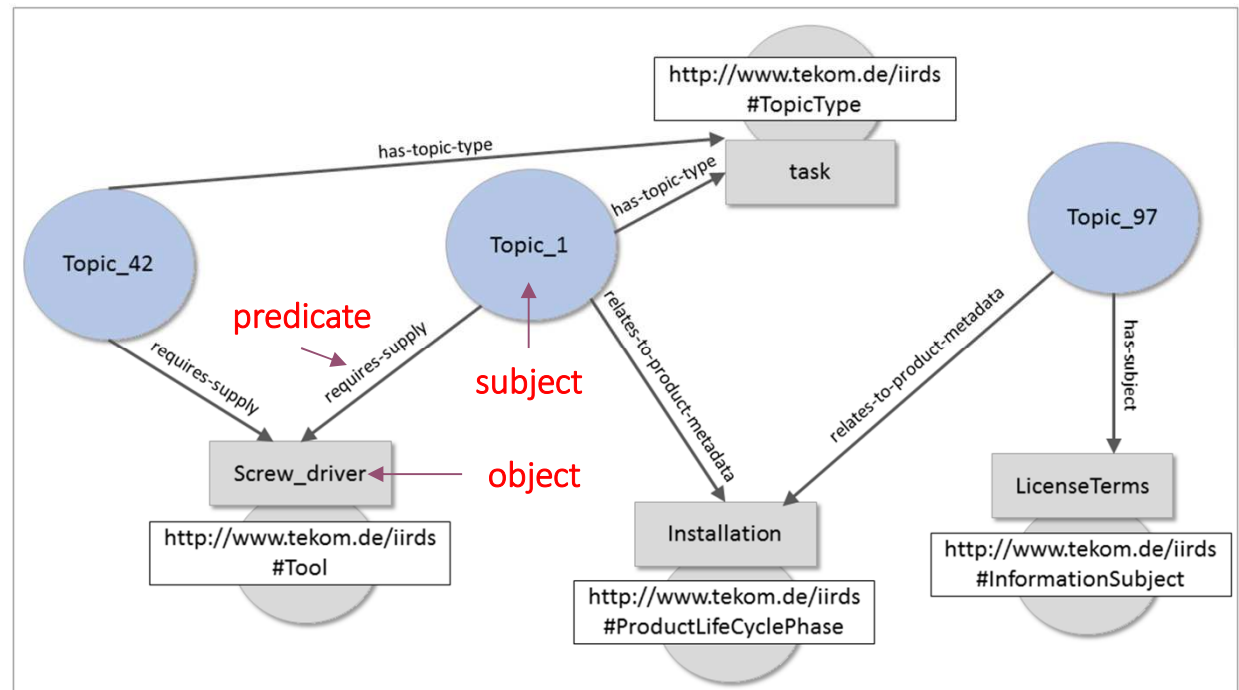


Image courtesy of parsons AG



Coming Soon: A Working Set of Plugins for iiRDS and DITA

Developed by parson and Empolis, set for release after iiRDS 1.0 is released, and will consist of the following two parts:

- A plugin for specialized DITA content, which makes the iiRDS metadata (PI classification) available via specialized attributes. The classification values can be assigned via subject scheme maps.
- A transformation plugin (based on HTML 5 transformation from DITA-OT) that generates iiRDS packages from DITA maps



DITA Code Samples with iIRDS

- With the parson DITA plugin, a new @qualification-role attribute is added to the DITA framework. In this example, the role is assigned to a table row:

```
<row information-  
subject="diagnostics"  
qualification-  
role="serviceTechnician">  
    <entry>The rotor does not  
turn even though the device is  
switched on.</entry>  
    <entry>The gear unit is  
damaged.</entry> <entry>Have the  
gear unit replaced.</entry>  
</row>
```

- Subject scheme maps provide the values for the attribute (customer-specific values in this case):

```
<subjectScheme>  
<subjectHead><subjectHeadMeta><navtitle>Skill  
level</navtitle></subjectHeadMeta>  
</subjectHead>  
<subjectdef keys="qualification-roleClass">  
    <subjectdef keys="service-technician">  
        <topicmeta><navtitle>Service  
technician</navtitle></topicmeta>  
    </subjectdef>  
    <subjectdef keys="operator">  
<topicmeta><navtitle>Operator</navtitle></top  
icmeta>  
    </subjectdef>  
</subjectdef>  
</subjectScheme>
```



DITA Code Samples with iiRDS

- Subject scheme maps provide the values for the attribute (customer-specific values in this case):
- Empolis plugin generates (in addition to the HTML files with the content) the RDF file with the metadata of the package. Metadata for this informationUnit:

```
<subjectScheme>
<subjectHead><subjectHeadMeta><navtitle>Skill
level</navtitle></subjectHeadMeta>
</subjectHead>
<subjectdef keys="qualification-roleClass">
  <subjectdef keys="service-technician">
    <topicmeta><navtitle>Service
technician</navtitle></topicmeta>
  </subjectdef>
  <subjectdef keys="operator">
<topicmeta><navtitle>Operator</navtitle></top
icmeta>
  </subjectdef>
</subjectdef>
</subjectScheme>
```

```
<iirds:requires-qualification>
  <iirds:Role>
    <rdfs:label>Service
technician</rdfs:label>
  </iirds:Role>
</iirds:requires-qualification>
```



So Where is iiRDS Today?

- Version 1.0 is set to launch later this month (April 2018)
- Integrates feedback from the RfC phase and harmonizes iiRDS with international standards on content/document management in industrial settings (e.g. IEC 82045-2 and IEC 61355-1)
- parson + Empolis plan to update the corresponding plugins for generating iiRDS from DITA
- Ralf Robers, the president of the iiRDS Consortium has said: "the iiRDS consortium wants to work and coordinate efforts with the DITA Technical Committee..."
- iiRDS Consortium (iirds.tekom.de/iirds-consortium/) is keen to have new members from the international community

A hand is shown in the upper right corner, holding a single wooden block. Below it, a row of five wooden blocks is laid out horizontally. The background is a soft-focus green and white, suggesting an outdoor setting. The text is centered on an orange banner that spans the width of the image.

Building with DITA: Chatbots and DITA



About Chatbots and AI

- While a “sexy” topic at recent techcomm conferences, much of this is FUD:
 - They are *not* likely to replace technical writer jobs (though call-center staff should be worried)
 - Little understanding of the business fundamentals and scenarios in which they work best
 - More of a focus on its supposed potential than the mechanics
 - Consensus is that it is too early in development for immediate, widespread use





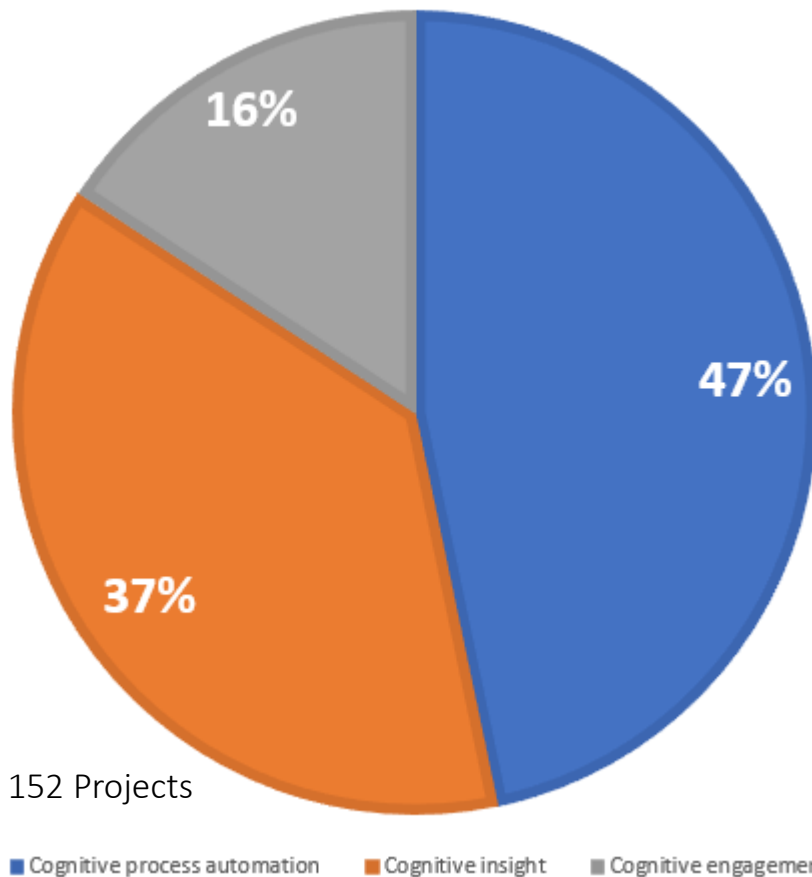
Business Capabilities and the Three Types of AI

- **Cognitive process automation:** typically back-office administrative and financial activities; e.g. “reading” legal and contractual documents using natural language processing
- **Cognitive insight:** detecting patterns in vast volumes of data and interpret their meaning; predict what a customer is likely to buy and pitch targeted ads
- **Cognitive engagement:** intelligent agents that offer 24/7 customer service addressing a broad and growing array of issues from password requests to technical support questions—all in the customer’s natural language



Harvard Business School Survey on AI Usage

COGNITIVE PROJECT DEPLOYMENTS BY TYPE



Count: 152 Projects

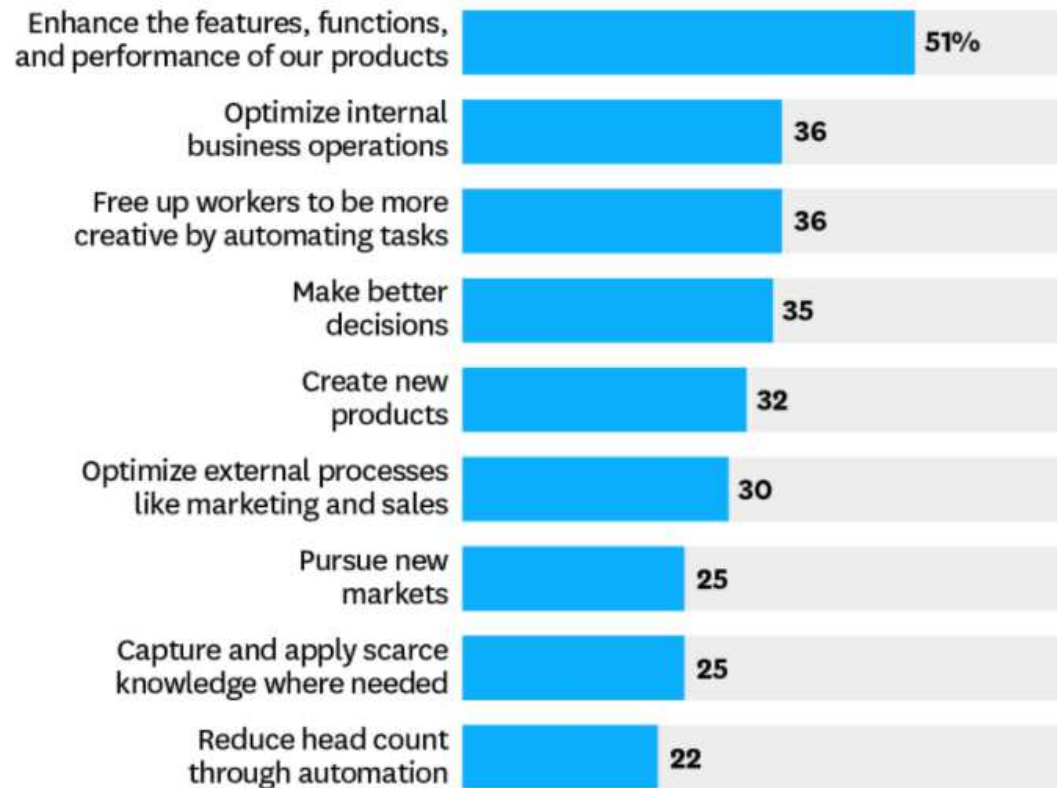
- “The companies in our study tended to use cognitive engagement technologies more to interact with employees than with customers. That may change as firms become more comfortable turning customer interactions over to machines.”
- Point being is that cognitive engagement chatbots are still very much an emerging technology, still considered too new within the business community for wide deployment *for now*.



Harvard Business School Survey on AI Usage

- Main intent of AI from a business perspective is to enhance brand/customer experience
- “Companies tend to take a conservative approach to customer-facing cognitive engagement technologies largely because of their immaturity.”

PERCENTAGE OF EXECUTIVES WHO CITE THE FOLLOWING AS BENEFITS OF AI





How Does DITA Fit In?

Chatbots take one of two paths when working with external content (including technical content):

1. Digests existing material which it incorporates directly within itself, or
 2. It refers to appropriate material a user wants via metadata
- Either way, having tech docs in a structured format (like DITA) makes it easier for chatbots to digest or find content to reference because well-defined content types + descriptive metadata make it easier for chatbots to use

Robot graphic
courtesy of
Vanessa Roberts



Integrating DITA Content with Chatbots

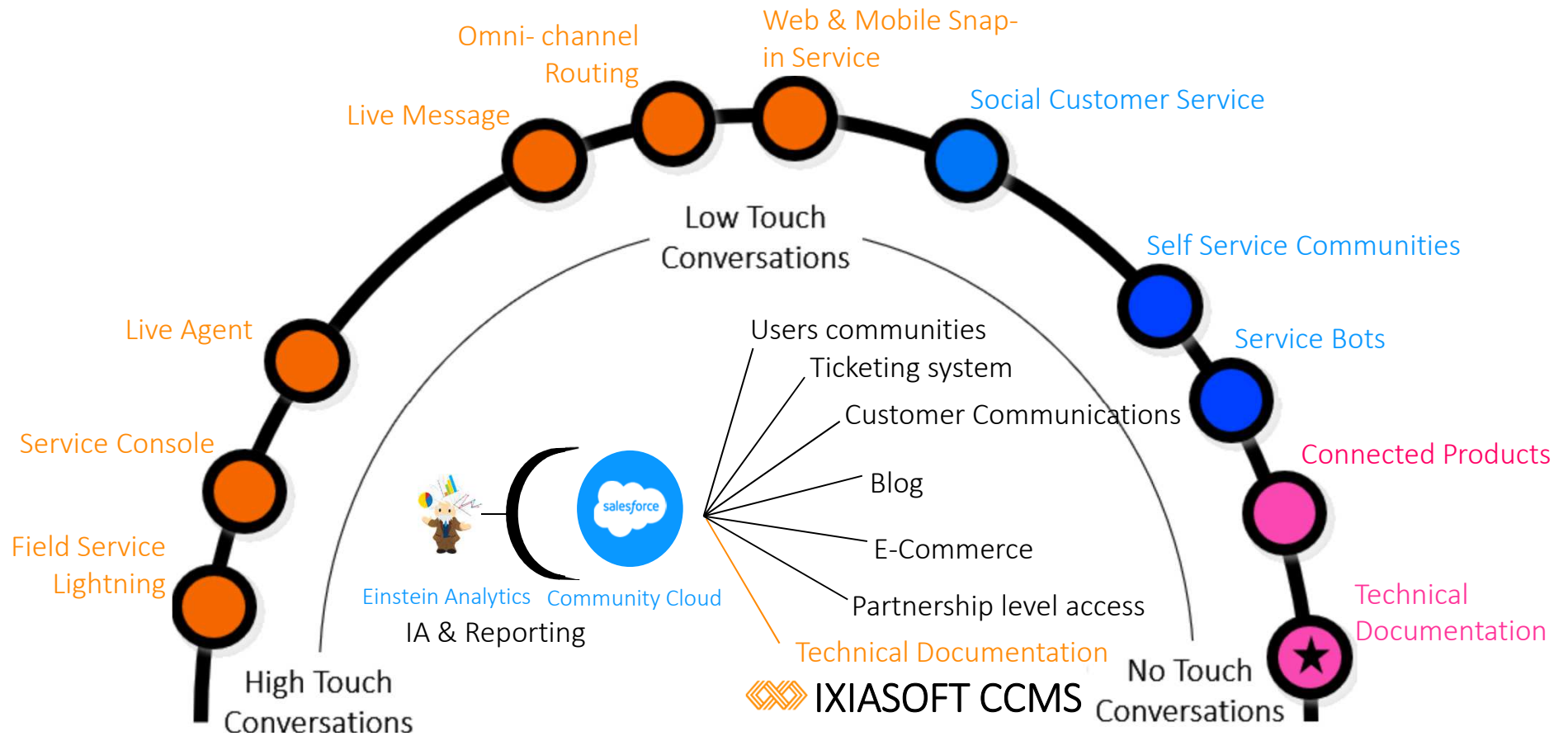
- Think of chatbots as a new, complementary channel for content delivery that works best with structured content
- DITA definitely has a role here; not directly in the chatbot “conversation”, but in providing technical content when it is needed



Image from
SoftBank Robotics US



Chatbots as *Part* of Overall User Experience





Are Chatbots Transformative?

- Potentially yes, but we are definitely in the *early* stages
- In the short term, *from a tech docs perspective*, it is likely to be rolled out as a complementary platform for answering users' queries
- From a business perspective, costs of failure are high; expect firms to be cautious when implementing chatbots



Image from icraftcorner.com

A hand is shown in the upper right corner, holding a single wooden block. Below it, a row of five wooden blocks is laid out horizontally. The background is a soft-focus green field under a bright sky. An orange banner with white text is positioned in the middle of the image.

DITA as a Building Block: Conclusions



There Are Now More Reasons for Using DITA

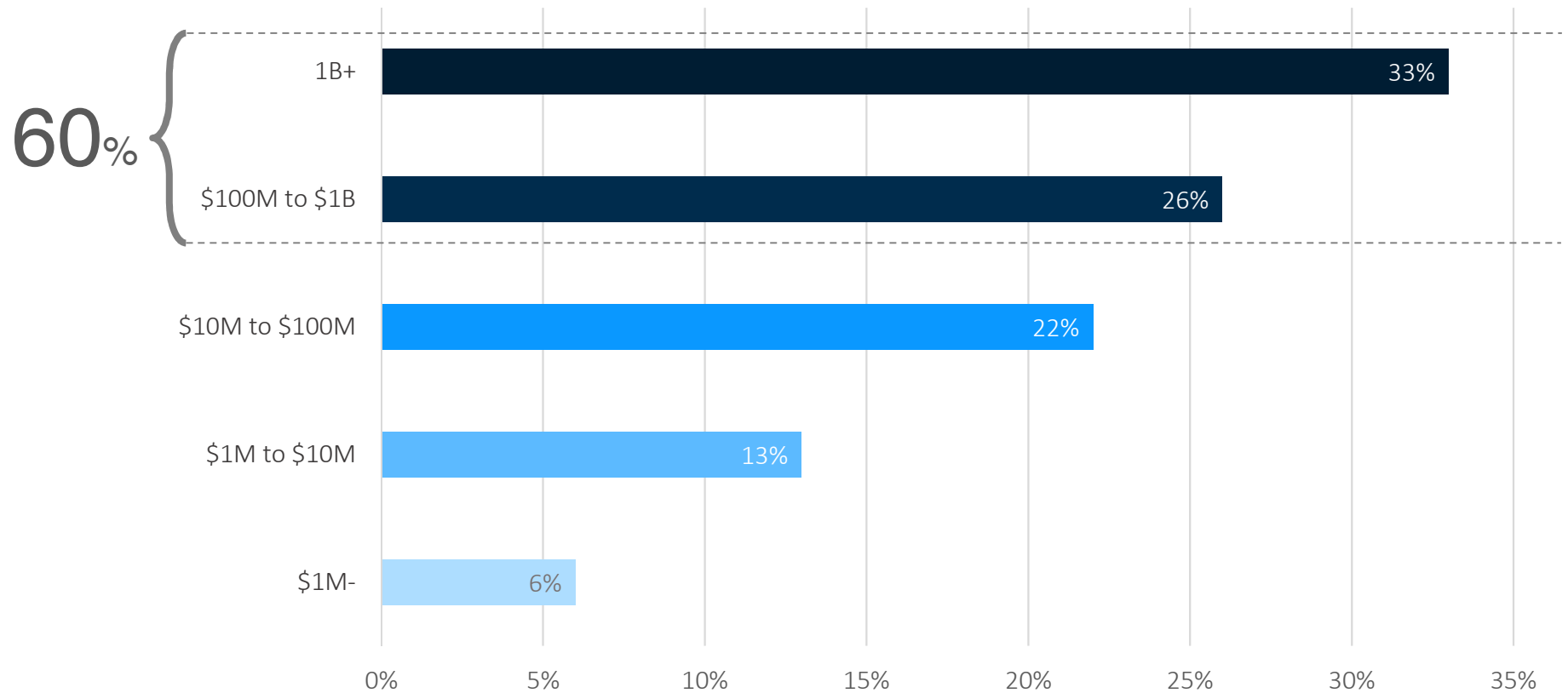
In addition to content reuse, reduced localization costs and other efficiencies, firms are beginning to see DITA as a building block for:

- Communicating with SMEs via LwDITA/MDITA
- Working with Industry 4.0 initiatives
- Effective chatbot development
- And this is probably only the start of what is possible!



Final Thought: Most DITA-using Firms Are Large

Sizes of Firms Using DITA by Annual Revenue



Sources: Hoovers, LinkedIn • Feb 2018 • 633 firms

- And therefore are most capable of gaining these further advantages of using DITA



Q&A

- Blog on www.ixiasoft.com
- Twitter: @IXIASOFT and @KeithIXIASOFT
- IXIASOFT DITA CMS Users LinkedIn group
- OASIS DITA Adoption Committee articles
- Member of OASIS DITA Technical Committee



IXIASOFT