COMMUNICATION STUDIES. СОЦІАЛЬНІ КОМУНІКАЦІЇ

УДК 378.147:070(477)

<https://doi.org/10.31548/philolog2020.01.096>

АНАЛІЗ СЕМАНТИЧНИХ ОДИНИЦЬ ТЕКСТОВОГО ФОРМАТУ NITF

О.Ю. БАЛАЛАЄВА, кандидат педагогічних наук, доцент,
Національний університет біоресурсів і природокористування України
E-mail: olena.balalaeva@gmail.com
<https://orcid.org/0000-0002-2675-5554>

Анотація. У статті розглянуто стандарт Міжнародної ради з питань преси та телекомунікацій щодо визначення змісту і структури статей та відповідних метаданих із використанням розширюваної мови розмітки XML для обміну новинним контентом. Мета статті – проаналізувати семантичні одиниці стандартизованого текстового формату NITF (News Industry Text Format). Наведено характеристики основних елементів контенту: PERSON – власні назви осіб, FUNCTION – статус, роль або функція особи, ORG – назва організації, LOCATION – місцезнаходження, EVENT – подія, OBJECT.TITLE – назва об'єкта, CHRON – позначення часу, MONEY – позначення грошових одиниць, NUM – числові дані у вигляді цифр або символів, Q та BQ – цитована інформація. Сьогодні стандарт NITF – це найпоширеніший словник XML серед видавців новин у всьому світі. Його використовують The New York Times, Deutsche Presse-Agentur, Agence France Presse, Nachrichtenagentur, Agenzia Nazionale Stampa Associata, Norwegian News Agency, CCI Europe тощо. Вивчення досвіду використання стандартів провідними інформаційними агентствами є перспективним для вітчизняної журналістики як в теоретичному, так і в прикладному аспектах.

Ключові слова: текстовий формат, семантична одиниця, контент, теґ, медіа.

Актуальність. Запровадження міжнародних стандартів професійної діяльності є актуальною проблемою для сучасної української журналістики. Як зазначає В. Шинкарук, в умовах інтеграції України до світової спільноти актуалізується потреба суспільства у висококваліфікованих журналістах, які володіють високою гуманітарною культурою і спеціальними знаннями, мають креативне мислення, мотивовані до сталого професійного розвитку, здатні коректно та об'єктивно опрацьовувати, інтерпретувати й поширювати інформацію [2, с. 121]. В. Різун зауважує, що «сьогодні медіагалузь потребує фахівців, всебічно обізнаних, готових працювати оперативним, з вільним використанням цифрової техніки... Натомість доводиться спостерігати прояви недостатнього професійного рівня журналістів, редакторів, рекламистів.

Як контент, так і форма подачі матеріалів виявляють їхню поверхову обізнаність у нових комунікаційних явищах, а також правових, стандартизаційних та інших аспектах функціонування новітніх медіа» [3, с. 5]. Українські журналісти нової формації мають орієнтуватися на досвід провідних світових інформаційних агентств і працювати з новими технічними стандартами, зокрема з чинними стандартами, розробленими Міжнародною радою з питань преси та телекомунікацій – IPTC (International Press Telecommunications Council) [1, с. 103].

Аналіз останніх досліджень та публікацій. Стандарт NITF розглядали закордонні дослідники Д. Ален, В. Бем [7], Д. Родрігез, Т. Хернандез [6], Т. Ягінума [8] та ін. В Україні питання впровадження міжнародних професійних стандартів

у практику журналістської діяльності висвітлено в працях Б. Потятиника, С. Квіта, В. Різуна [3], В. Шинкарука [2] та багатьох інших дослідників. Однак увагу в цих дослідженнях зацентовано переважно на етичних та соціальних, а не технічних аспектах проблеми.

Мета статті – проаналізувати семантичні одиниці стандартизованого текстового формату NITF (News Industry Text Format). Дослідження проведено на матеріалі стандарту NITF офіційного сайту IPTC [5] із застосуванням методів аналізу наукових та нормативних джерел, узагальнення, систематизації.

Результати дослідження. IPTC запустила проект текстового формату для новин на початку 1990-х, коли члени організації почали шукати наступника форматів ANPA 1312 і IPTC 7901, які забезпечували загальну платформу для новинних служб і газет для обміну контентом. У результаті був розроблений News Industry Text Format, чинна версія якого NITF 3.6 вийшла 2012 р. Сьогодні NITF – це відкритий, широко використовуваний, добре документований і підтримуваний стандарт для структурування новинних статей.

Текстовий формат NITF використовує розширювану мову розмітки XML для визначення змісту та структури новинних статей. Оскільки метадані можуть бути застосовані до всього новинного контенту, документи NITF набагато доступніші й зручніші для пошуку, ніж HTML-сторінки. Використовуючи NITF, видавці можуть адаптувати зовнішній вигляд, взаємодію та інтерактивність документів до пропускнуї спроможності, пристроїв та індивідуальних потреб клієнтів. Ці документи можна переконвертувати в HTML, WML (для бездротових пристроїв), RTF (для друку) або в будь-який інший формат, що його побажає видавець [5].

Семантичними одиницями, або елементами контенту в NITF є такі: PERSON, FUNCTION, ORG, LOCATION, EVENT, OBJECT.TITLE, CHRON, MONEY, NUM, Q та BQ.

Найважливішими, на думку дослідни-

ків стандарту, є три проблеми: правильна та послідовна класифікація семантичних одиниць, належний ступінь специфічності, диференціації та деталізації, а також відношення між елементами контенту, які перебувають у безпосередній близькості в тексті [7, с. 1]. Вміст тегів є символічними даними, відмежованими від тексту початковими та кінцевими тегами певного елемента контенту. Визначення контенту тегів відповідає на питання, що саме ми хочемо згрупувати за певною смисловою категорією.

Під тегом PERSON 'особа' вказують власні назви переважно диференційовані на імена та прізвища. Наприклад: `<PERSON><NAME.GIVEN>Volodymyr</NAME.GIVEN
<NAME.FAMILY>Zelensky</NAME.FAMILY></PERSON>`.

Під цим тегом вказують також ввічливі форми звертання (*Mr, Mrs, Frau etc.*), титули академічні (*Dr., Prof. etc.*) і дворянські (*Queen, Dame, Sir, Lord, Lady etc.*). Часто згадуване в новинному повідомленні ім'я людини передуює або супроводжується згадуванням статусу чи ролі, пов'язаних із цією людиною (наприклад, *Ганна Новосад, міністр освіти і науки*). NITF дозволяє зробити цей показник явним, помістивши тег FUNCTION до тегу PERSON.

Тег FUNCTION 'функція' указує статус чи роль персони і, як правило, зазначає повні офіційні титули або звання (політичні, церковні, військові, цивільні). Наприклад, *секретар соціального забезпечення, виконавчий директор* тощо. Він також кодує синоніми й варіанти, які використовують у пресі для офіційних повних назв, наприклад, *Президент* замість *президент Сполучених Штатів Америки*.

Основний концепт (заголовок або ядро фрази) можна розширити шляхом премодифікації та постмодифікації. Кожна така модифікація збільшує специфікацію концепту. Наприклад, базовий концепт *президент* може бути розширений таким чином: *новообраний президент* (премодифікація), *президент України* (постмодифікація),

колишній президент України (пре- і пост-модифікація).

Розробники стандарту рекомендують для підвищення точності обирати мінімально можливий рівень абстракції, щоб теґи охоплювали найконкретніший вираз, що трапляється в тексті. При цьому модифікатори мають зовсім іншу природу. Вони можуть бути потрібні для однозначного посилання або передавати надлишкову, додаткову інформацію для користувача [7, с. 2].

Теґ FUNCTION використовують також для ідентифікації представників професії або посади: *юрист, журналіст, детектив, тренер, актор тощо*, а також для позначення родинних стосунків: *батько, мати, чоловік, дружина, брат тощо*.

Теґ ORG 'організація' ідентифікує назву організації – групи людей, об'єднаних спільною метою та певними правилами співпраці – з будь-якої сфери діяльності: політичної, громадської, комерційної, військової, спортивної тощо. Наприклад, *Національна спілка журналістів України, Фізкультурно-спортивне товариство «Динамо», Банк Ангелії, Singapore Telecommunications Ltd, Bundesanstalt für Angestellte, FC Bayern München*. Модифікаторам іменника назви організації може бути власна назва міста або країни з відсиланням на місце, де організація знаходиться або національність, до якої вона належить.

Теґ LOCATION 'місцезнаходження' ідентифікує географічне розташування чи важливі місця й містить простий текст або структуровану інформацію. Підкатегорії СУБЛОКАЦІЯ, МІСТО, РЕГІОН, ШТАТ, КРАЇНА можуть використовувати в послідовної комбінації для географічних назв, у порядку зростання за їхнім розміром або адміністративним значенням, наприклад, для диференціації назв об'єктів (*Миколаїв, Львівська область і Миколаїв, Миколаївська область*). Теґ LOCATION застосовують і для географічних об'єктів (річки, озера, гори, моря, острови тощо), і для відомих штучних споруд (будівлі, споруди, мости, аеропорти, автостради тощо).

Назви місцезнаходження можуть мати різні функції в новинній статті: вони можуть бути самостійними семантичними одиницями чи модифікаторами для інших об'єктів, таких як PERSON, FUNCTION, ORG та EVENT.

Теґ EVENT 'подія' має бути обмежений новинними подіями. Факторами цінності новин є значимість, близькість, популярність залучених осіб, наслідок, екстраординарність, людський інтерес, актуальність тощо. Події можуть бути диференційовані на типи подій (*Загальні вибори, Олімпійські ігри, Flugzeugabsturz*) і на конкретні випадки із зазначенням місця та часу (*Вибори президента України 2019, Літні Олімпійські ігри 2020; der TWA-Absturz*).

Під теґом CHRON 'час' указують точні дати, а також дні тижня, релігійні свята, позначення відносного часу, які можуть бути приписані конкретній даті, та часові інтервали. Дата вставляється як атрибут у форматі YYYYMMDD. Наприклад: 20200222 – 22 лютого 2020 року.

Теґ MONEY 'гроші' маркує назви грошових одиниць (валюту, тип одиниці) та кількість, згадувані в новинному повідомленні.

Теґом NUM 'числові дані' позначають усі числа у вигляді цифр або символів, за винятком елементів MONEY і CHRON. Елемент NUM може містити текст, числа, структуровану інформацію.

Для цитованої інформації використовують теґи Q – коротка цитата в тексті та BQ – довга цитата в тексті, блок цитування.

Для отримання зворотного зв'язку та пропозицій щодо змін у стандарті, а також для обміну досвідом та обговорення дискусійних питань на офіційному сайті Міжнародної ради з питань преси та телекомунікацій створено форум користувачів NITF [5]. Група є відкритою для всіх, хто зацікавлений у розробленні та оптимізації цього стандарту.

Висновки. Сьогодні стандарт News Industry Text Format (NITF) – це найпоширеніший словник XML серед видавців новин у всьому світі. Його

семантичні одиниці, належний ступінь специфічності, диференціації та деталізації дозволяють чітко визначати зміст і структуру новинної статті та відповідних метаданих, завдяки чому цей стандарт використовують *The New York Times*, *Deutsche Presse-Agentur*, *Agence*

France Presse, *Agenzia Nazionale Stampa Associata*, *CCI Europe* та ін. Вивчення досвіду використання стандартів провідними інформаційними агентствами є перспективним для вітчизняної журналістики як в теоретичному, так і в прикладному аспектах.

Список використаних джерел

1. Балалаєва О. Ю. Аналіз новинних кодів за стандартами Міжнародної ради з питань преси та телекомунікацій. *Науковий вісник Національного університету біоресурсів і природокористування України. Серія: Філологічні науки*. 2017. Вип. 272. С. 103–110.

2. Шинкарук В. Д., Балалаєва О. Ю. Підготовка аграрних журналістів: перспективи та завдання. *Науковий вісник Національного університету біоресурсів і природокористування України. Серія: Філологічні науки*. 2017. Вип. 263. С. 117–125.

3. Новітні медіа та комунікаційні технології: комплекс навч. програм для спец. «Журналістика», «Видавнича справа та редагування», «Реклама та зв'язки з громадськістю» / За ред. В. Шевченко. Київ: Паливода, 2012. 412 с.

4. Balalaieva O. Media topics in descriptors of International Press Telecommunications Council. *Вісник Львівського університету. Серія Журналістика*. 2017. № 42. С. 151–158.

5. IPTC: Official Website. URL: <http://www.iptc.org> (Last accessed: 02.02.2020).

6. Hernandez T., Rodriguez, D. Usos documentales del mercado de texto periodístico: NewsML y NITF. *Contenidos y Aspectos Legales en la Sociedad de la Información*. Valencia: Universitat Politècnica de Valencia, 2002. URL: <http://hdl.handle.net/10016/893> (Last accessed: 03.01.2020).

7. Möhr W., Allen D. Considerations for the Semantic Markup with the NITF. URL: <https://www.iptc.org/std/NITF/documentation/stx9804-NITFmarkupGuidelines.pdf> (Last accessed: 29.01.2020).

8. Yaginuma T., Pereira T., Baptista A.

Design of Metadata Elements for Digital News Articles in the Omnipaper Project. *From information to knowledge: 7th ICC/IFIP International Conference on Electronic Publishing, ELPUB*. Minho, Portugal, June 25–28, 2003. P. 132–139.

References

1. Balalaieva, O. Yu. (2017). Analiz novynnykh kodiv za standartamy Mizhnarodnoi rady z pytan presy ta telekomunikatsii [Analysis of new codes under the standards of the International Press Telecommunications Council]. *Naukovyi visnyk Natsionalnoho universytetu bioresursiv i pryrodokorystuvannia Ukrainy. Seriya: Filohichni nauky. (Scientific Herald of National University of Life and Environmental Sciences of Ukraine. Series: Philological Sciences)*. Issue 272. 103–110.

2. Shynkaruk, V. D., Balalaieva, O. Yu. (2017). Pidhotovka ahrarnykh zhurnalistiv: perspektyvy ta zavdannia [Training of agrarian journalists: prospects and goals]. *Naukovyi visnyk Natsionalnoho universytetu bioresursiv i pryrodokorystuvannia Ukrainy. Seriya: Filohichni nauky. (Scientific Herald of National University of Life and Environmental Sciences of Ukraine. Series: Philological Sciences)*. Issue 263, 117–125.

3. Shevchenko, V. ed. (2012). Novitni media ta komunikatsiini tekhnolohii: kompleks navch. prohram dlia spets. «Zhurnalistyka», «Vydavnycha справа ta redahuvannia», «Reklama ta zviazky z hromadskistiu» [New media and communication technologies: curricula for the specialties "Journalism", "Publishing and editing", "Advertising & Public Relations"]. Kyiv: Palyvoda, 412.

4. Balalaieva, O. Yu. (2017). Media topics in descriptors of International Press Telecommunications Council. *Visnyk Lvivskoho natsionalnoho universytetu. Seriya*

Zhurnalistyka (Herald of the Lviv University. Series: Journalism). Issue 42, 151–158.

5. IPTC: Official Website. URL: <http://www.iptc.org>.

6. Hernández, T., Rodríguez, D. (2002). Usos documentales del marcado de texto periodístico: NEWSML y NITF. Contenidos y Aspectos Legales en la Sociedad de la Información. Valencia: Universitat Politècnica de Valencia. URL: <http://hdl.handle.net/10016/893>.

7. Möhr, W., Allen, D. (1998).

Considerations for the Semantic Markup with the NITF. URL: <https://www.iptc.org/std/NITF/documentation/stx9804-NITFmarkupGuidelines.pdf>.

8. Yaginuma, T., Pereira T., Baptista, A. (2003). Design of Metadata Elements for Digital News Articles in the Omnipaper Project. From information to knowledge: 7th ICC/IFIP International Conference on Electronic Publishing, ELPUB. Minho, Portugal, June 25-28, 2003, 132–139.

ANALYSIS OF SEMANTIC UNITS OF NITF TEXT FORMAT

O. Yu. Balalaieva

Abstract. *The article deals with the International Press Telecommunications Council open standard NITF on determining the content and structure of news articles and related metadata using XML, designed for sharing content and news interchange.*

The implementation of international standards of professional activity is a current issue for Ukrainian journalism. Modern researchers point out a lack of professional competence of journalists, editors, advertisers, and others; both the content and the form of a material presentation show authors superficial knowledge of new communication phenomena, as well as legal, standardization and other aspects of the new media. A new generation of Ukrainian journalists should be guided by the experience of world-leading news agencies and work with new technical standards.

The purpose of the article is to analyze semantic units of standardized News Industry Text Format (NITF). The research is based on the material of the NITF standard from the official International Press Telecommunications Council site using the methods of analysis of scientific and normative sources, generalization and systematization.

Some characteristics of the main content elements or semantic units specified in the News Industry Text Format are outlined. These include: PERSON 'personal names', FUNCTION 'status, role or function of the person, family relations', ORG 'name of organization', LOCATION 'indication of place, location', EVENT 'anything that happens, news-relevant event', OBJECT.TITLE 'name of news-relevant object', CHRON 'expression of time', MONEY 'monetary items, currency names, NUM 'numerical data, except for the expression of time (concrete dates, time intervals) and money items (amount of money)', Q and BQ 'quoted information (inline quotation and block quotation)'.

NITF's rational semantic units, multiple categorization, the proper degree of specificity, detailing and differentiation make it possible to clearly define the content and structure of the news article and relevant metadata, that is why this standard is used by such newswires as The New York Times, Deutsche Presse-Agentur, Agence France Presse, Agenzia Nazionale Stampa Associata, presstext Nachrichtenagentur, Norwegian News Agency, CCI Europe and others. Today, NITF is the most widely used XML dictionary among news publishers worldwide. Studying the experience of using standards by leading news agencies is promising for Ukrainian journalism in both theoretical and applied aspects.

Key words: *text format, semantic unit, content, tags, media.*